Forum of Mathematics, Pi (2014), Vol. 2, e1, 56 pages
doi:10.1017/fmp.2014.1

THE BREUIL–MÉZARD CONJECTURE FOR POTENTIALLY BARSOTTI–TATE REPRESENTATIONS

TOBY GEE[1] and MARK KISIN[2]

[1] Department of Mathematics, Imperial College London SW7 2RH, UK;
email: toby.gee@imperial.ac.uk

[2] Department of Mathematics, Harvard University, Cambridge, MA 02138, USA;
email: kisin@math.harvard.edu

Received 18 September 2013; accepted 9 October 2014

Abstract

We prove the Breuil–Mézard conjecture for two-dimensional potentially Barsotti–Tate representations of the absolute Galois group G_K, K a finite extension of $\mathbb{Q}_p$, for any $p > 2$ (up to the question of determining precise values for the multiplicities that occur). In the case that $K/\mathbb{Q}_p$ is unramified, we also determine most of the multiplicities. We then apply these results to the weight part of Serre's conjecture, proving a variety of results including the Buzzard–Diamond–Jarvis conjecture.

2010 Mathematics Subject Classification: 11F33

Overview

The Breuil–Mézard conjecture [**BM02**] predicts the Hilbert–Samuel multiplicity of the special fibre of a deformation ring of a mod p local Galois representation. Its motivation is to give a local explanation for the multiplicities seen in Hecke algebras and spaces of modular forms; in that sense it can be viewed as an avatar for the hoped for p-adic local Langlands correspondence. For $\mathrm{GL}_2(\mathbb{Q}_p)$, the conjecture was mostly proved in [**Kis09a**], using global methods and the p-adic local Langlands correspondence, and it was used in that paper to deduce modularity lifting theorems.

The conjecture was originally formulated for two-dimensional representations of $G_{\mathbb{Q}_p}$ (the absolute Galois group of $\mathbb{Q}_p$) with a restriction on the Hodge–Tate weights of the deformations under consideration, but the formulation extends immediately to the case of unrestricted regular Hodge–Tate weights. In fact, there is a natural generalization of the Breuil–Mézard conjecture for continuous representations $\bar{r} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ for any finite extension $K/\mathbb{Q}_p$; see [Kis10], and Section 2 below. (In fact, it is possible to formulate a generalization for representations of arbitrary dimension; see [EG14, Section 4].)

There is currently no known generalization of the p-adic local Langlands correspondence to $\mathrm{GL}_2(K)$, $K \neq \mathbb{Q}_p$, and it is accordingly not possible to use the local methods of [Kis09a] to prove the conjecture in greater generality. The main idea of the present paper is that one can *use* the modularity lifting theorems proved in [Kis09b] and [Gee06] (by a completely different method, unrelated to p-adic local Langlands) to prove the Breuil–Mézard conjecture for all potentially Barsotti–Tate deformation rings. As a byproduct of these arguments, we are also able to prove the Buzzard–Diamond–Jarvis conjecture [BDJ10] on the weight part of Serre's conjecture for Hilbert modular forms, as well as its generalizations to arbitrary totally real fields conjectured in [Sch08] and [Gee11a].

1. Introduction

Fix finite extensions $K/\mathbb{Q}_p$ and $E/\mathbb{Q}_p$, the latter (which will be our coefficient field) assumed sufficiently large. Let E have ring of integers $\mathcal{O}$, uniformizer π, and residue field $\mathbb{F}$, let k be the residue field of K, and fix a continuous representation $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$. Given a Hodge type λ and an inertial type τ (see Section 2 below for the precise definitions of these notions, and of the other objects recalled without definition in this introduction), there is a universal lifting $\mathcal{O}$-algebra $R_{\bar{r}}^{\square,\lambda,\tau}$ for potentially crystalline lifts of $\bar{r}$ of Hodge type λ and Galois type τ. The Breuil–Mézard conjecture predicts the Hilbert–Samuel multiplicity $e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi)$ in terms of the representation theory of $\mathrm{GL}_2(\mathcal{O}_K)$. More precisely, a result of Henniart attaches to τ a smooth, irreducible, finite-dimensional E-representation $\sigma(\tau)$ of $\mathrm{GL}_2(\mathcal{O}_K)$ via the local Langlands correspondence, and there is also an algebraic representation W_λ of $\mathrm{GL}_2(\mathcal{O}_K)$ associated to λ. Let $L_{\lambda,\tau} \subset W_\lambda \otimes \sigma(\tau)$ be a $\mathrm{GL}_2(\mathcal{O}_K)$-invariant lattice; then the general shape of the Breuil–Mézard conjecture is that for all λ, τ we have

$$e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi) = \sum_{\sigma} n_{\lambda,\tau}(\sigma)\mu_\sigma(\bar{r}),$$

where σ runs over the irreducible mod p representations of $\mathrm{GL}_2(k)$, $n_{\lambda,\tau}(\sigma)$ is the multiplicity of σ as a Jordan–Hölder factor of $L_{\lambda,\tau}/\pi$, and the $\mu_\sigma(\bar{r})$ are nonnegative integers, depending only on $\bar{r}$ and σ (and not on λ or τ).

One can view the conjecture as giving infinitely many equations (corresponding to the different possibilities for λ and τ) in the finitely many unknowns $\mu_\sigma(\bar{r})$, and it is easy to see that, if the conjecture is completely proved, then the $\mu_\sigma(\bar{r})$ are completely determined (in fact, they are determined by the equations for τ the trivial representation and λ 'small', and are zero unless σ is a predicted Serre weight for $\bar{r}$ in the sense of [Gee11a]). Our main theorem (see Corollary 4.5.6) is the following result, which establishes the conjecture in the potentially Barsotti–Tate case ($\lambda = 0$ in the terminology above).

THEOREM A. *Suppose that $p > 2$. Then there are uniquely determined nonnegative integers $\mu_\sigma(\bar{r})$ such that, for all inertial types τ, we have*

$$e(R_{\bar{r}}^{\square,0,\tau}/\pi) = \sum_\sigma n_{0,\tau}(\sigma)\mu_\sigma(\bar{r}).$$

Furthermore, the $\mu_\sigma(\bar{r})$ enjoy the following properties.

(1) *$\mu_\sigma(\bar{r}) \neq 0$ if and only if σ is a predicted Serre weight for $\bar{r}$.*

(2) *If $K/\mathbb{Q}_p$ is unramified and σ is regular, then $\mu_\sigma(\bar{r}) = e(R_{\bar{r}}^{\square,\sigma}/\pi)$, where $R_{\bar{r}}^{\square,\sigma}$ is the crystalline lifting ring of Hodge type determined by σ. If furthermore σ is Fontaine–Laffaille regular, then $\mu_\sigma(\bar{r}) = 1$ if σ is a predicted Serre weight for $\bar{r}$, and is zero otherwise.*

See the introduction to [**GLS13**] for a discussion of the history of the various definitions of predicted Serre weights for $\bar{r}$ and of their equivalence, and see Section 2.2 below for the precise definitions we are using. We are able to apply this result and the techniques that we use to prove it to the problem of the weight part of Serre's conjecture. For any $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ as above, we define $W^{\mathrm{BT}}(\bar{r})$ to be the set of weights σ such that $\mu_\sigma(\bar{r}) > 0$. It follows from Theorem A(1) that $W^{\mathrm{BT}}(\bar{r})$ is precisely the set of weights predicted by the Buzzard–Diamond–Jarvis conjecture and its generalizations [**BDJ10**, **Sch08**, **Gee11a**]. We then prove the following result (see Corollary 5.5.4).

THEOREM B. *Let $p > 2$ be prime, let F be a totally real field, and let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Assume that $\overline{\rho}$ is modular, that $\overline{\rho}|_{G_{F(\zeta_p)}}$ is irreducible, and if $p = 5$ assume further that the projective image of $\overline{\rho}|_{G_{F(\zeta_p)}}$ is not isomorphic to A_5.*

For each place $v \mid p$ of F with residue field k_v, let σ_v be a Serre weight of $\mathrm{GL}_2(k_v)$. Then $\overline{\rho}$ is modular of weight $\otimes_{v|p}\sigma_v$ if and only if $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all v.

In the case that p is unramified in F, this proves the Buzzard–Diamond–Jarvis (BDJ) conjecture [**BDJ10**] for $\overline{\rho}$. More generally, by Theorem A(1) (which relies on the main result of [**GLS13**]), it proves the generalizations of the BDJ conjecture to arbitrary totally real fields conjectured in [**Sch08**] and [**Gee11a**]. In particular, it shows that the set of weights for which $\overline{\rho}$ is modular depends only on the restrictions of $\overline{\rho}$ to decomposition groups at places dividing p, which was not previously known. We remark that, in the case of indefinite quaternion algebras, another proof of the BDJ conjecture (which, like the proof given in this paper, relies on the version of the conjecture for unitary groups proven in [**BLGG13b**, **GLS14**, **GLS13**]) is given in [**New13**].

We emphasize that, in the above theorem, the definition of $\overline{\rho}$ being modular of some weight is in terms of quaternion algebras as in the Buzzard–Diamond–Jarvis conjecture (see [**BDJ10**, Definition 2.1, Conjecture 3.14] and [**Gee11b**]), rather than in terms of unitary groups as in [**BLGG13b**]. The former statement is more subtle, since the set of *local* Serre weights in the Buzzard–Diamond–Jarvis conjecture is attached to $\bar{r}$ by considering crystalline liftings, while, for a quaternion algebra, one cannot lift every global Serre weight to characteristic zero.

We now describe some of the techniques of this paper in more detail. In Sections 3 and 4, we adapt the patching arguments of [**Kis09a**] to the case of a rank two unitary group over a totally real field F^+. In the presence of a modularity lifting theorem, these yield a relationship between a patched Hecke module and a tensor product of the local deformation rings which appear in the Breuil–Mézard conjecture. In particular, the modularity lifting theorems proved in [**Kis09b**] and [**Gee06**] imply that such a relationship holds in the (two-dimensional) potentially Barsotti–Tate case. More generally, we show that such a relationship holds for representations which are 'potential diagonalizable' in the sense of [**BLGGT14a**].

These arguments produce not a solution in nonnegative integers to the systems of equations that we seek, but rather a solution to a product, over the primes $\mathfrak{p}|p$ of F^+, of these systems of equations. We deduce that a solution for each individual system exists by showing that such a solution is unique, and applying some linear algebra. (It does not seem possible to avoid dealing with a product of systems of equations by, for example, choosing F^+ so that p is inert in F^+: our methods require that we realize the local representation $\bar{r}$ globally. We do this via the potential automorphy techniques of [**BLGGT14a**] and [**Cal12**] (see Appendix A), and these methods cannot ensure that p is inert in F^+. (In the case that $\bar{r}$ is irreducible or decomposable, it is presumably possible to avoid this by making use of CM forms, but they cannot handle the case that $\bar{r}$ is reducible and indecomposable.)) It follows from the construction that $\mu_\sigma(\bar{r}) \neq 0$ if and only if there are modular forms of weight σ for the unitary group used in the construction. Using this, together with the results on Serre's conjecture for unitary groups proved in [**BLGG13b**, **GLS14**, **GLS13**], we deduce the other claims in Theorem A.

In order to prove Theorem B and the Buzzard–Diamond–Jarvis conjecture, we repeat these constructions in Section 4, in the setting of the cohomology of

Shimura curves associated to division algebras. The uniqueness of the $\mu_\sigma(\bar{r})$ implies that the multiplicities computed globally in this setting agree with those computed via unitary groups. On the other hand, as for unitary groups, by construction these multiplicities are nonzero if and only if there are modular forms of weight σ for the quaternion algebra used in the construction. Theorem B follows from this, and implies the BDJ conjecture and its generalizations when combined with Theorem A(1).

The argument comparing multiplicities for unitary groups and quaternion algebras seems to us to be analogous to the use of the trace formula to prove instances of functoriality for inner forms. We view the left-hand side of the equality

$$e(R_{\bar{r}}^{\square,0,\tau}/\pi) = \sum_{\sigma} n_{0,\tau}(\sigma)\mu_\sigma(\bar{r})$$

as being the 'geometric' side, and the right-hand side as the 'spectral' side. Then the geometric side is manifestly the same in the unitary group or Shimura curve settings, from which we deduce that the spectral sides are also the same, and thus transfer the proof of the Buzzard–Diamond–Jarvis conjecture from the unitary group context to the original setting of [**BDJ10**]. As mentioned above, the proof of Theorem A(1) uses the results of [**BLGG13b**, **GLS14**, **GLS13**] which, in turn, rely crucially on the fact that there is no parity restriction on the weights of the modular forms on unitary groups. Thus, it does not seem possible to prove Theorem A(1) and its consequences for the BDJ conjecture by working only with quaternion algebras.

1.1. Notation. Fix an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$, and an algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ for each prime p. Fix also an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ for each p. If M is a finite extension of $\mathbb{Q}$ or $\mathbb{Q}_p$, we let G_M denote its absolute Galois group. If M is a finite extension of $\mathbb{Q}_p$ for some p, we write I_M for the inertia subgroup of G_M. If F is a number field and v is a finite place of F, then we let Frob_v denote a geometric Frobenius element of G_{F_v}.

If R is a local ring, we write $\mathfrak{m}_R$ for the maximal ideal of R. We write all matrix transposes on the left; so ${}^tA$ is the transpose of A.

Let ε denote the p-adic cyclotomic character, and $\bar{\varepsilon} = \omega$ the mod p cyclotomic character.

If K is a finite extension of $\mathbb{Q}_p$ for some p, we let rec 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 $\mathrm{rec}(\pi)$ is a complex Weil–Deligne representation of the Weil group W_K. Choose an isomorphism $\imath : \overline{\mathbb{Q}}_p \xrightarrow{\sim} \mathbb{C}$, and set $r_p(\pi) := \imath^{-1} \circ \mathrm{rec} \circ \imath(\pi \otimes |\det|^{(1-n)/2})$; this is independent of the choice of $\imath$. We let $\mathrm{Art}_K : K^\times \to W_K^{ab}$

be the isomorphism provided by local class field theory, which we normalize so that uniformizers correspond to geometric Frobenius elements. We will sometimes identify characters of I_K and of $\mathcal{O}_K^\times$ via Art_K without comment. If F is a number field, we write Art_F for the global Artin map, normalized to be compatible with the Art_{F_v}.

2. Potentially crystalline deformation rings

In this section, we recall the formulation of the Breuil–Mézard conjecture, or rather its generalization to finite extensions of $\mathbb{Q}_p$ (see [Kis10]).

We fix a finite extension $K/\mathbb{Q}_p$ with ring of integers $\mathcal{O}_K$ and residue field k. Let $E \subset \overline{\mathbb{Q}}_p$ be a finite extension of $\mathbb{Q}_p$ with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$. In particular, we may regard $\mathbb{F}$ as a subfield of $\overline{\mathbb{F}}_p$, the residue field of $\overline{\mathbb{Q}}_p$. We assume throughout the paper that E is sufficiently large; in particular, we assume that E contains the image of every embedding $K \hookrightarrow \overline{\mathbb{Q}}_p$, and that various $\overline{\mathbb{Q}}_p$-representations $\tau, \sigma(\tau)$ that we consider are in fact defined over E.

2.1. The Breuil–Mézard conjecture. Let B be a finite local E-algebra, and let V_B be a finite free B-module, with a continuous potentially semistable action of G_K. Then

$$D_{\mathrm{dR}}(V_B) = (B_{\mathrm{dR}} \otimes_{\mathbb{Q}_p} V)^{G_K}$$

is a filtered $B \otimes_{\mathbb{Q}_p} K$-module which is free of rank $\mathrm{rk}_B V_B$, and whose associated graded is projective over $B \otimes_{\mathbb{Q}_p} K$. For an embedding $\varsigma : K \hookrightarrow E$, we denote by $\mathrm{HT}_\varsigma(V_B)$ the multiset in which the integer i appears with multiplicity $\mathrm{rk}_B \mathrm{gr}^i(D_{\mathrm{dR}}(V_B) \otimes_{B \otimes_{\mathbb{Q}_p} K, \varsigma} B)$. We call the elements of $\mathrm{HT}_\varsigma(V_B)$ the Hodge–Tate numbers of V_B with respect to ς. Thus, for example, $\mathrm{HT}_\varsigma(\varepsilon) = \{-1\}$.

Let $\mathbb{Z}^2_+$ denote the set of pairs (λ_1, λ_2) of integers with $\lambda_1 \geqslant \lambda_2$. Fix $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$, and suppose that $\mathrm{rk}_B V_B = 2$. We say that V_B has *Hodge type* λ if, for each $\varsigma : K \hookrightarrow E$, the Hodge–Tate weights of V_B with respect to ς are $\lambda_{\varsigma,1} + 1$ and $\lambda_{\varsigma,2}$. If V_B is potentially crystalline and has Hodge type zero, we say that V_B is *potentially Barsotti–Tate*. (Note that this is a slight abuse of terminology; however, we will have no reason to deal with representations with nonregular Hodge–Tate weights, and so we exclude them from consideration.) (Note that it is more usual in the literature to say that V_B is potentially Barsotti–Tate if it is potentially crystalline, and $V_B^\vee$ has Hodge type zero; in particular, all of our definitions are dual to those of [BM02], but of course our results can be translated into their setting by taking duals.)

An *inertial type* is a representation $\tau : I_K \to \mathrm{GL}_2(E)$ with open kernel, with the property that (possibly after replacing E with a finite extension) τ may be extended to a representation of the Weil group W_K. In particular, it is semisimple and factors through a finite quotient of I_K. We say that V_B is of *Galois type* τ if the traces of elements of I_K acting on $\mathrm{D}_{\mathrm{pst}}(V_B)$ and τ are equal.

Let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. Let $R_{\bar{r}}^{\square}$ be the universal $\mathcal{O}$-lifting ring of $\bar{r}$, so that $R_{\bar{r}}^{\square}$ pro-represents the functor which assigns to a local Artin $\mathcal{O}$-algebra R with residue field $\mathbb{F}$ the set of liftings of $\bar{r}$ to a representation $G_K \to \mathrm{GL}_2(R)$.

The following is a special case of one of the main results of [Kis08].

PROPOSITION 2.1.1. *There is a unique (possibly zero) p-torsion free quotient* $R_{\bar{r}}^{\square,\lambda,\tau}$ *of* $R_{\bar{r}}^{\square}$ *such that, for any finite local* E*-algebra* B*, and any* E*-homomorphism* $x : R_{\bar{r}}^{\square} \to B$*, the* B*-representation of* G_K *induced by* x *is potentially crystalline of Galois type* τ *and Hodge type* λ*, if and only if* x *factors through* $R_{\bar{r}}^{\square,\lambda,\tau}$.

Moreover, $\mathrm{Spec}\, R_{\bar{r}}^{\square,\lambda,\tau}[1/p]$ *is formally smooth and everywhere of dimension* $4 + [K : \mathbb{Q}_p]$.

In the case that τ is the trivial representation, we will drop it from the notation, and write $R_{\bar{r}}^{\square,\lambda}$ for $R^{\square,\lambda,\tau}$. We will need the following simple lemma later.

LEMMA 2.1.2. *Let* $\overline{\psi} : G_K \to \mathbb{F}^{\times}$ *be an unramified character. Then the* $\mathcal{O}$*-algebras* $R_{\bar{r}}^{\square,\lambda,\tau}$ *and* $R_{\bar{r}\otimes(\overline{\psi}\circ\det)}^{\square,\lambda,\tau}$ *are isomorphic.*

Proof. Let ψ denote the Teichmüller lift of $\overline{\psi}$; since this character is unramified, it is crystalline with all Hodge–Tate weights equal to zero. If $r^{\mathrm{univ}} : G_K \to \mathrm{GL}_2(R_{\bar{r}}^{\square,\lambda,\tau})$ is the universal lift of $\bar{r}$, then $r^{\mathrm{univ}} \otimes (\psi \circ \det) : G_K \to \mathrm{GL}_2(R_{\bar{r}}^{\square,\lambda,\tau})$ is the universal lift of $\bar{r} \otimes (\overline{\psi} \circ \det)$, as required. □

Let $\tau : I_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be an inertial type, as above. We have the following theorem of Henniart (see the appendix to [BM02]).

THEOREM 2.1.3. *There is a finite-dimensional irreducible* $\overline{\mathbb{Q}}_p$*-representation* $\sigma(\tau)$ *of* $\mathrm{GL}_2(\mathcal{O}_K)$ *such that, for any two-dimensional Frobenius semisimple representation* $\tilde{\tau}$ *of* WD_K*,* $(r_p^{-1}(\tilde{\tau})^{\vee})|_{\mathrm{GL}_2(\mathcal{O}_K)}$ *contains* $\sigma(\tau)$ *if and only if* $\tilde{\tau}|_{I_K} \sim \tau$ *and* $N = 0$ *on* $\tilde{\tau}$*. Furthermore, for all* $\tilde{\tau}$*, we have*

$$\dim_{\overline{\mathbb{Q}}_p} \mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_K)}(\sigma(\tau), r_p^{-1}(\tilde{\tau})^{\vee}) \leqslant 1.$$

If $|k| > 2$*, then* $\sigma(\tau)$ *is uniquely determined.*

2.1.4. For $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$, we set

$$W_\lambda = \otimes_\varsigma \left(\left(\det{}^{\lambda_{\varsigma,2}} \otimes \mathrm{Sym}^{\lambda_{\varsigma,1}-\lambda_{\varsigma,2}} \mathcal{O}_K^2 \right) \otimes_{\mathcal{O}_{K,\varsigma}} \mathcal{O} \right),$$

where ς runs over the embeddings $K \hookrightarrow E$.

Now assume that E is sufficiently large that $\sigma(\tau)$ is defined over E. Let π be a uniformizer of $\mathcal{O}$. Fix a $\mathrm{GL}_2(\mathcal{O}_K)$-stable $\mathcal{O}$-lattice $L_{\lambda,\tau} \subset W_\lambda \otimes_{\mathcal{O}} \sigma(\tau)$. If A is a Noetherian local ring, we let $e(A)$ denote the Hilbert–Samuel multiplicity of A.

CONJECTURE 2.1.5 (Breuil–Mézard). There exist nonnegative integers $\mu_\sigma(\bar{r})$ for each irreducible mod p representation σ of $\mathrm{GL}_2(k)$ such that, for any τ, λ as above, we have

$$e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi) = \sum_\sigma n(\sigma)\mu_\sigma(\bar{r}),$$

where σ runs over isomorphism classes of irreducible mod p representations of $\mathrm{GL}_2(k)$, and $n(\sigma) = n_{\lambda,\tau}(\sigma)$ is the multiplicity of σ as a Jordan–Hölder factor of $L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F}$, so that

$$(L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{ss} \xrightarrow{\sim} \oplus_\sigma \sigma^{n(\sigma)}.$$

REMARK 2.1.6. When $|k| = 2$, so that the type $\sigma(\tau)$ is not necessarily unique, it follows from [BD13, Proposition 4.2] that the quantities $n(\sigma)$ are independent of the choice of $\sigma(\tau)$.

2.2. Local Serre weights.

The following definitions will be useful to us later, in order to give more explicit information about the $\mu_\sigma(\bar{r})$ in some cases.

2.2.1. By a *(local) Serre weight* we mean an absolutely irreducible representation of $\mathrm{GL}_2(k)$ on an $\mathbb{F}$-vector space, up to isomorphism. Let $(\mathbb{Z}^2_+)^{\mathrm{Hom}\,(k,\mathbb{F})}_{\mathrm{sw}} \subset (\mathbb{Z}^2_+)^{\mathrm{Hom}\,(k,\mathbb{F})}$ be the subset consisting of elements a such that

$$p - 1 \geqslant a_{\varsigma,1} - a_{\varsigma,2}$$

for each $\varsigma \in \mathrm{Hom}\,(k, \mathbb{F})$. Then

$$\sigma_a = \otimes_\varsigma \det{}^{a_{\varsigma,2}} \otimes \mathrm{Sym}^{a_{\varsigma,1}-a_{\varsigma,2}} k^2 \otimes_{k,\varsigma} \mathbb{F},$$

where ς runs over the embeddings $k \hookrightarrow \mathbb{F}$, is a Serre weight, and every Serre weight is of this form. We say that $a, a' \in (\mathbb{Z}^2_+)^{\mathrm{Hom}\,(k,\mathbb{F})}_{\mathrm{sw}}$ are *equivalent* if $\sigma_a \cong \sigma_{a'}$. If $\sigma_a \cong \sigma_{a'}$, then $a_{\varsigma,1} - a_{\varsigma,2} = a'_{\varsigma,1} - a'_{\varsigma,2}$ for all ς.

2.2.2. We have a natural surjection $\mathrm{Hom}_{\mathbb{Q}_p}(K,E) \twoheadrightarrow \mathrm{Hom}(k,\mathbb{F})$, which is a bijection if and only if $K/\mathbb{Q}_p$ is unramified. Suppose that $K/\mathbb{Q}_p$ has ramification degree e. For each $\varsigma \in \mathrm{Hom}(k,\mathbb{F})$, we choose an element $\tau_{\varsigma,1}$ in the preimage of ς, and denote the remaining elements of the preimage by $\tau_{\varsigma,2},\dots,\tau_{\varsigma,e}$. Now, given $a \in (\mathbb{Z}^2_+)^{\mathrm{Hom}(k,\mathbb{F})}$, we define $\lambda_a \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$ as follows. For $i = 1, 2$, $\lambda_{a_{\tau_{\varsigma,1},i}} = a_{\varsigma,i}$, and $\lambda_{a_{\tau_{\varsigma,j},i}} = 0$ if $j > 1$. When $K/\mathbb{Q}_p$ is unramified, we will sometimes write a for λ_a.

By definition, we have $W_{\lambda_a} \otimes_{\mathcal{O}} \mathbb{F} \cong \sigma_a$, and we write $R^{\square,a}$ for $R^{\square,\lambda_a}$. Note that in the case that $K/\mathbb{Q}_p$ is ramified this definition depends on the choice of the places $\tau_{\varsigma,1}$, as does Definition 2.2.3 below (at least *a priori*). However, our only use of this definition in the ramified case will be to prove that in certain cases $R^{\square,\lambda_a}$ is nonzero (see Lemma 4.3.10 below), and in these cases our argument will in fact show that this holds for any choice of the $\tau_{\varsigma,1}$.

If $a, a' \in (\mathbb{Z}^2_+)^{\mathrm{Hom}(k,\mathbb{F})}_{\mathrm{sw}}$ with $\sigma_a \cong \sigma_{a'}$, then there is a crystalline character $\psi_{a,a'}$ of G_K with trivial reduction mod p and Hodge–Tate weights given by $\mathrm{HT}_{\tau_{\varsigma,1}}(\psi_{a,a'}) = a_{\varsigma,2} - a'_{\varsigma,2}$, and $\mathrm{HT}_{\tau_{\varsigma,j}}(\psi_{a,a'}) = 0$ if $j > 1$. The corresponding universal deformation to $R^{\square,a}$ is obtained from that to $R^{\square,a'}$ by twisting by $\psi_{a,a'}$, which induces an isomorphism $R^{\square,a} \cong R^{\square,a'}$.

If σ is an irreducible $\mathbb{F}$-representation of $\mathrm{GL}_2(k)$ and $\sigma \cong \sigma_a$, we will write $R^{\square,\sigma}$ for $R^{\square,a}$. The following definitions will be needed in order to state our main results.

Definition 2.2.3. We say that $\bar{r}$ has a *lift of Hodge type* σ if (with notation as above) $R^{\square,\sigma} \neq 0$.

Definition 2.2.4. Suppose that $K/\mathbb{Q}_p$ is unramified. We say that a Serre weight σ is *regular* if $\sigma \cong \sigma_a$ for some $a \in (\mathbb{Z}^2_+)^{\mathrm{Hom}(k,\mathbb{F})}$, with $p-2 \geqslant a_{\varsigma,1} - a_{\varsigma,2}$ for each $\varsigma \in \mathrm{Hom}(k,\mathbb{F})$. We say that it is *Fontaine–Laffaille regular* if for each $\varsigma \in \mathrm{Hom}(k,\mathbb{F})$ we have $p-3 \geqslant a_{\varsigma,1} - a_{\varsigma,2}$. The remarks above show that these conditions depend only on σ and not on the choice of a.

2.2.5. We remark that this is a significantly less restrictive definition than the definition of regular weights in [**Gee11b**].

In order to make use of the results of [**GLS12**, **GLS14**, **GLS13**] we need to recall the notion of a predicted Serre weight. Beginning with the seminal work of [**BDJ10**], various definitions have been formulated of conjectural sets of Serre weights for two-dimensional global mod p representations (see [**Sch08**, **Gee11a**]). These sets are defined purely locally, and the relationship between the different local definitions is important in proving the weight part of Serre's conjecture; see [**BLGG13b**, Section 4] for a thorough discussion. In particular, given a continuous representation $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$, sets of weights $W^{\mathrm{explicit}}(\bar{r})$ and $W^{\mathrm{cris}}(\bar{r})$ are defined in [BLGG13b]. The set $W^{\mathrm{explicit}}(\bar{r})$ is defined by an explicit recipe that generalizes those of [BDJ10, Sch08] and [Gee11a], whereas the set $W^{\mathrm{cris}}(\bar{r})$ is the set of weights σ for which $\bar{r}$ has a crystalline lift of Hodge type σ. It is shown in [BLGG13b] that $W^{\mathrm{explicit}}(\bar{r}) \subset W^{\mathrm{cris}}(\bar{r})$, and conjectured that equality holds; this has now been proved [GLS13]. We make the following definition.

DEFINITION 2.2.6. If $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ is a continuous representation, we say that σ is a *predicted Serre weight* for $\bar{r}$ if $\sigma \in W^{\mathrm{explicit}}(\bar{r})$, where $W^{\mathrm{explicit}}(\bar{r})$ is the set of Serre weights defined in [BLGG13b, Section 4].

3. Algebraic automorphic forms and Galois representations

3.1. Unitary groups and algebraic automorphic forms. Let $p > 2$ be a prime, and let F be an imaginary CM field with maximal totally real field subfield F^+. We assume throughout this paper that the following hold.

- F/F^+ is unramified at all finite places.
- Every place $v \mid p$ of F^+ splits in F.

By class field theory, the set of places v of F^+ such that -1 is not in the image of the local norm map $(F \otimes_{F^+} F_v^+)^\times \to F_v^{+,\times}$ has even cardinality. Since we are assuming that F/F^+ is unramified at all finite places, this implies that $[F^+ : \mathbb{Q}]$ is even.

3.1.1. We now define the unitary groups with which we shall work. It will be convenient to define these as groups over $\mathcal{O}_{F^+}$.

Let $c \in \mathrm{Gal}\,(F/F^+)$ be the complex conjugation. The map $g \mapsto ({}^t g^c)^{-1}$ is an involution of $\mathrm{GL}_{2/\mathcal{O}_F}$ which covers the action of c on $\mathcal{O}_F$. Since $\mathcal{O}_F$ is unramified over $\mathcal{O}_{F^+}$, it follows by étale descent that there is a reductive group G over $\mathcal{O}_{F^+}$ such that, for any $\mathcal{O}_{F^+}$-algebra R, one has

$$G(R) = \{g \in \mathrm{GL}_2(\mathcal{O}_F \otimes_{\mathcal{O}_{F^+}} R) : {}^t g^c g = 1\}.$$

Thus G is a unitary group which is definite at infinite places, and which is quasisplit at finite places, because -1 is in the image of the local norm map at finite places. (G is automatically quasisplit if n is odd, and if n is even this is equivalent to the standard Hermitian form being a sum of hyperbolic planes, by (for example) the discussion of groups of type ${}^2A_n$ on [Tit66, p. 55]. If

$N(a) = -1$, then the equation $z_1 z_1^c + z_2 z_2^c = 0$ has solution vectors of the form (z, az) and $(z, a^c z)$, so the condition is satisfied for n even.)

By construction, G is equipped with an isomorphism $\iota : G_{/\mathcal{O}_F} \xrightarrow{\sim} \mathrm{GL}_{2/\mathcal{O}_F}$ such that $\iota \circ c \circ \iota^{-1}(g) = ({}^t g^c)^{-1}$. If v is a place of F^+ which splits as ww^c over F, then ι induces isomorphisms

$$\iota_w : G(\mathcal{O}_{F_v^+}) \xrightarrow{\sim} \mathrm{GL}_2(\mathcal{O}_{F_w}), \iota_{w^c} : G(\mathcal{O}_{F_v^+}) \xrightarrow{\sim} \mathrm{GL}_2(\mathcal{O}_{F_{w^c}})$$

which satisfy $\iota_w \circ \iota_{w^c}^{-1}(g) = ({}^t g^c)^{-1}$. These extend to isomorphisms $\iota_w : G(F_v^+) \xrightarrow{\sim} \mathrm{GL}_2(F_w)$ and $\iota_{w^c} : G(F_v^+) \xrightarrow{\sim} \mathrm{GL}_2(F_{w^c})$ with the same properties.

3.1.2. Continue to let $E/\mathbb{Q}_p$ be a sufficiently large extension with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$. Let S_p denote the set of places of F^+ lying over p, and for each $v \in S_p$ fix a place $\widetilde{v}$ of F lying over v. Let $\tilde{S}_p$ denote the set of places $\tilde{v}$ for $v \in S_p$. Write $F_p^+ = F^+ \otimes_{\mathbb{Q}} \mathbb{Q}_p$, and $\mathcal{O}_{F_p^+}$ for the p-adic completion of $\mathcal{O}_{F^+}$.

Let W be an $\mathcal{O}$-module with an action of $G(\mathcal{O}_{F_p^+})$, and let $U \subset G(\mathbb{A}_{F^+}^\infty)$ be a compact open subgroup with the property that, for each $u \in U$, if u_p denotes the projection of u to $G(F_p^+)$, then $u_p \in G(\mathcal{O}_{F_p^+})$.

Let $S(U, W)$ denote the space of algebraic modular forms on G of level U and weight W, that is, the space of functions

$$f : G(F^+)\backslash G(\mathbb{A}_{F^+}^\infty) \to W$$

with $f(gu) = u_p^{-1} f(g)$ for all $u \in U$.

For any compact open subgroup U of $G(\mathbb{A}_{F^+}^\infty)$ as above, we may write $G(\mathbb{A}_{F^+}^\infty) = \coprod_i G(F^+) t_i U$ for some finite set $\{t_i\}$. Then there is an isomorphism

$$S(U, W) \xrightarrow{\sim} \oplus_i W^{U \cap t_i^{-1} G(F^+) t_i}$$

given by $f \mapsto (f(t_i))_i$. We say that U is *sufficiently small* if for some finite place v of F^+ the projection of U to $G(F_v^+)$ contains no element of finite order other than the identity. Suppose that U is sufficiently small. Then for each i as above we have $U \cap t_i^{-1} G(F^+) t_i = \{1\}$, so we see that, for any $\mathcal{O}$-algebra A and $\mathcal{O}$-module W, we have

$$S(U, W \otimes_{\mathcal{O}} A) \cong S(U, W) \otimes_{\mathcal{O}} A.$$

3.1.3. Let $\widetilde{I}_p$ denote the set of embeddings $F \hookrightarrow E$ giving rise to a place in $\widetilde{S}_p$. For any $v \in S_p$, let $\widetilde{I}_{\widetilde{v}}$ denote the set of elements of $\widetilde{I}_p$ lying over $\widetilde{v}$. Note that $|\widetilde{I}_{\widetilde{v}}| = [F_v^+ : \mathbb{Q}_p] = [F_{\widetilde{v}} : \mathbb{Q}_p]$. Let $\mathbb{Z}_+^2$ be as in Section 2. For any $\lambda \in (\mathbb{Z}_+^2)^{\widetilde{I}_{\widetilde{v}}}$, let

W_λ be the E-vector space with an action of $\mathrm{GL}_2(\mathcal{O}_{F_{\tilde{v}}})$ given by

$$W_\lambda := \otimes_{\varsigma \in \widetilde{I}_{\tilde{v}}} \det{}^{\lambda_{\varsigma,2}} \otimes \mathrm{Sym}^{\lambda_{\varsigma,1}-\lambda_{\varsigma,2}} F_{\tilde{v}}^2 \otimes_{F_{\tilde{v}},\varsigma} E.$$

We give this an action of $G(\mathcal{O}_{F_v^+})$ via $\iota_{\tilde{v}}$.

For any $\lambda \in (\mathbb{Z}_+^2)^{\widetilde{I}_p}$ and $v \in S_p$, let $\lambda_v \in (\mathbb{Z}_+^2)^{\widetilde{I}_{\tilde{v}}}$ denote the tuple of pairs in λ indexed by $\widetilde{I}_{\tilde{v}}$, and let W_λ be the E-vector space with an action of $G(\mathcal{O}_{F_p^+})$ given by

$$W_\lambda := \otimes_{v \in S_p} W_{\lambda_v}.$$

For each $v \in S_p$, let τ_v be an inertial type for $G_{F_v^+}$, so that (since E is assumed sufficiently large) there is an absolutely irreducible E-representation $\sigma(\tau_v)$ of $\mathrm{GL}_2(\mathcal{O}_{F_{\tilde{v}}})$ associated to τ_v by Theorem 2.1.3. Write $\sigma(\tau)$ for the tensor product of the $\sigma(\tau_v)$, regarded as a representation of $G(\mathcal{O}_{F_p^+})$ by letting $G(\mathcal{O}_{F_p^+})$ act on $\sigma(\tau_v)$ via $\iota_{\tilde{v}}$. Fix a $G(\mathcal{O}_{F_p^+})$-stable $\mathcal{O}$-lattice $L_{\lambda,\tau} \subset W_\lambda \otimes_{\mathcal{O}} \sigma(\tau)$, and, for any $\mathcal{O}$-module A, write

$$S_{\lambda,\tau}(U, A) := S(U, L_{\lambda,\tau} \otimes_{\mathcal{O}} A).$$

3.2. Hecke algebras and Galois representations

DEFINITION 3.2.1. We say that a compact open subgroup of $G(\mathbb{A}_{F^+}^\infty)$ is *good* if $U = \prod_v U_v$ with U_v a compact open subgroup of $G(F_v^+)$ such that the following hold.

- $U_v \subset G(\mathcal{O}_{F_v^+})$ for all v which split in F.
- $U_v = G(\mathcal{O}_{F_v^+})$ if $v \mid p$ or v is inert in F.

3.2.2. Let U be a good compact open subgroup of $G(\mathbb{A}_{F^+}^\infty)$. Let T be a finite set of finite places of F^+ which split in F, containing S_p and all the places v which split in F for which $U_v \neq G(\mathcal{O}_{F_v^+})$. We let $\mathbb{T}^{T,\mathrm{univ}}$ be the commutative $\mathcal{O}$-polynomial algebra generated by formal variables $T_w^{(j)}$ for $j = 1, 2$, and w a place of F lying over a place v of F^+ which splits in F and is not contained in T. For any $\lambda \in (\mathbb{Z}_+^2)^{\widetilde{I}_p}$, the algebra $\mathbb{T}^{T,\mathrm{univ}}$ acts on $S_{\lambda,\tau}(U, \mathcal{O})$ via the Hecke operators

$$T_w^{(j)} := \iota_w^{-1} \left[GL_2(\mathcal{O}_{F_w}) \begin{pmatrix} \varpi_w 1_j & 0 \\ 0 & 1_{2-j} \end{pmatrix} GL_2(\mathcal{O}_{F_w}) \right]$$

for $w \notin T$ and ϖ_w a uniformizer in $\mathcal{O}_{F_w}$.

We denote by $\mathbb{T}_{\lambda,\tau}^T(U, \mathcal{O})$ the image of $T^{T,\mathrm{univ}}$ in $\mathrm{End}_{\mathcal{O}}(S_{\lambda,\tau}(U, \mathcal{O}))$.

3.2.3. Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}^{T,\mathrm{univ}}$ with residue field $\mathbb{F}$. We say that $\mathfrak{m}$ is *automorphic* if $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \neq 0$ for some (λ,τ) as above. If $\bar{r} : G_F \to \mathrm{GL}_2(\mathbb{F})$ is an absolutely irreducible continuous representation, we say that $\bar{r}$ is *automorphic* if there are U, T as above and an automorphic maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{T,\mathrm{univ}}$ such that, for all places $v \notin T$ of F^+ which split as $v = ww^c$ in F, $\bar{r}|_{G_{F_w}}$ is unramified, and $\bar{r}(\mathrm{Frob}_w)$ has characteristic polynomial equal to the image of $X^2 - T_w^{(1)}X + (\mathbf{N}w)T_w^{(2)}$ in $\mathbb{F}[X]$. Note that, if $\bar{r}$ is automorphic, it is necessarily the case that $\bar{r}^c \cong \bar{r}^\vee\bar{\varepsilon}^{-1}$, because we have $T_{w^c}^{(j)} = (T_w^{(2)})^{-1}T_w^{(2-j)}$ in $\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})$.

Let $\mathcal{G}_2$ be the group scheme over $\mathbb{Z}$ defined to be the semidirect product of $\mathrm{GL}_2 \times \mathrm{GL}_1$ by the group $\{1, j\}$, which acts on $\mathrm{GL}_2 \times \mathrm{GL}_1$ by

$$j(g,\mu)j^{-1} = (\mu\, {}^t g^{-1}, \mu).$$

We have a homomorphism $\nu : \mathcal{G}_2 \to \mathrm{GL}_1$, sending (g,μ) to μ and j to -1.

Assume that $\bar{r}$ is absolutely irreducible and automorphic with corresponding maximal ideal $\mathfrak{m}$. By [CHT08, Lemma 2.1.4] and the main result of [BC11], we can and do extend $\bar{r}$ to a representation $\overline{\rho} : G_{F^+} \to \mathcal{G}_2(\mathbb{F})$ with $\nu \circ \overline{\rho} = \overline{\varepsilon}^{-1}$ and $\overline{\rho}|_{G_F} = (\bar{r}, \overline{\varepsilon}^{-1})$. By [CHT08, Lemma 2.1.4], the $\mathbb{F}^\times$-conjugacy classes of such extensions are a torsor under $\mathbb{F}^\times/(\mathbb{F}^\times)^2$. In particular, any two extensions $\overline{\rho}$ become conjugate if we replace $\mathbb{F}$ by a quadratic extension. We fix a choice of $\overline{\rho}$ from now on.

In the rest of the paper, we will make a number of arguments that will be vacuous unless $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \neq 0$ for the specific (λ,τ) at hand, but for technical reasons we do not assume this. Let $G_{F^+,T} := \mathrm{Gal}\,(F(T)/F^+)$, $G_{F,T} := \mathrm{Gal}\,(F(T)/F)$, where $F(T)$ is the maximal extension of F unramified outside of places lying over T.

THEOREM 3.2.4. *For any* (λ,τ) *there is a unique continuous lift*

$$\rho_{\mathfrak{m}} : G_{F^+,T} \to \mathcal{G}_2(\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}})$$

of $\overline{\rho}$*, which satisfies the following.*

(1) $\rho_{\mathfrak{m}}^{-1}((\mathrm{GL}_2 \times \mathrm{GL}_1)(\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}})) = G_{F,T}$.

(2) $\nu \circ \rho_{\mathfrak{m}} = \varepsilon^{-1}$.

(3) $\rho_{\mathfrak{m}}$ *is unramified outside* T*. If* $v \notin T$ *splits as* ww^c *in* F*, then* $\rho_{\mathfrak{m}}(\mathrm{Frob}_w)$ *has characteristic polynomial*

$$X^2 - T_w^{(1)}X + (\mathbf{N}w)T_w^{(2)}.$$

(4) *For each place* $v \in S_p$*, and each homomorphism* $x : \mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \to \overline{\mathbb{Q}}_p$*,* $x \circ \rho_{\mathfrak{m}}|_{G_{F_{\tilde{v}}}}$ *is potentially crystalline of Hodge type* λ_v *and Galois type* τ_v*.*

Proof. This may be proved in the same way as [**CHT08**, Proposition 3.4.4], making use of [**Lab11**, Corollaire 5.3], [**BLGGT14b**, Theorem 1.1], and Theorem 2.1.3 above. (More specifically, [**Lab11**, Corollaire 5.3] is used in order to transfer our automorphic forms to GL_n, in place of the arguments made in [**CHT08**, Proof of Proposition 3.3.2]. The argument of [**CHT08**, Proposition 3.4.4] then goes through unchanged, except that we have to check property (4) above; but, by Theorem 2.1.3, this is a consequence of local–global compatibility at places dividing p, which is a special case of [**BLGGT14b**, Theorem 1.1].) □

3.3. Global Serre weights. A *global Serre weight* (for G) is an absolutely irreducible mod p representation of $G(\mathcal{O}_{F_p^+})$ considered up to equivalence. For $v \in S_p$, denote by k_v the residue field of v. Let $a = (a_v)_{v\in S_p}$, where $a_v \in (\mathbb{Z}^2_+)^{\mathrm{Hom}\,(k_v,\mathbb{F})}_{\mathrm{sw}}$. We set

$$\sigma_a = \otimes_{\mathbb{F}}\sigma_{a_v},$$

where, for $v \in S_p$, σ_{a_v} is the representation of $\mathrm{GL}_2(k_v) = \mathrm{GL}_2(k_{\tilde{v}})$ defined in Section 2. We let $G(\mathcal{O}_{F_p^+})$ act on σ_{a_v} by the composite of $\iota_{\tilde{v}}$, and reduction modulo p. This makes σ_a an irreducible $\mathbb{F}$-representation of $G(\mathcal{O}_{F_p^+})$, and any irreducible $\mathbb{F}$-representation of $G(\mathcal{O}_{F_p^+})$ is equivalent to σ_a for some a.

We say that two such tuples $a = (a_v)_{v\in S_p}$, $a' = (a'_v)_{v\in S_p}$, are *equivalent* if $\sigma_a \cong \sigma_{a'}$; this implies that $a_{v,\varsigma,1} - a_{v,\varsigma,2} = a'_{v,\varsigma,1} - a'_{v,\varsigma,2}$ for each $v \in S_p$, and $\varsigma \in \mathrm{Hom}\,(k_v, \mathbb{F})$.

For $v \in S_p$, $\widetilde{I}_{\tilde{v}}$ naturally surjects onto $\mathrm{Hom}\,(k_v, \mathbb{F})$. Fixing once and for all a splitting of each of these surjections, we obtain, as in Section 2, a Hodge type $\lambda_{a_v} \in (\mathbb{Z}^2_+)^{\widetilde{I}_{\tilde{v}}}$, and hence an element $\lambda_a \in (\mathbb{Z}^2_+)^{\widetilde{I}_p}$.

4. The patching argument

4.1. Hecke algebras. In this section, we employ the Taylor–Wiles–Kisin patching method, following the approaches of [**Kis09a**] and [**BLGG11**] (which in turn follows [**CHT08**]). In particular, in the actual implementation of the patching method we follow [**BLGG11**] very closely.

4.1.1. Continue to assume that F is an imaginary CM field with maximal totally real field subfield F^+ such that the following hold.

- F/F^+ is unramified at all finite places.
- Every place $v \mid p$ of F^+ splits in F.
- $[F^+ : \mathbb{Q}]$ is even.

Let $G_{/\mathcal{O}_{F^+}}$ be the algebraic group defined in Section 3.

Fix an absolutely irreducible representation $\bar{r} : G_F \to \mathrm{GL}_2(\mathbb{F})$. Assume that the following hold.

- $\bar{r}$ is automorphic in the sense of Section 3.2 (so that, in particular, $\bar{r}^c \cong \bar{r}^\vee \bar{\varepsilon}^{-1}$).
- $\bar{r}$ is unramified at all primes $v \nmid p$.
- $\zeta_p \notin F$.
- $\bar{r}|_{G_{F(\zeta_p)}}$ is absolutely irreducible.
- The projective image of $\bar{r}$ is not isomorphic to A_4.

Let U be a good compact open subgroup of $G(\mathbb{A}^\infty_{F^+})$ (see Definition 3.2.1) such that, if $U_v \subset G(\mathcal{O}_{F_v})$ is not maximal for some $v \nmid p$, then the following hold.

- U_v is the preimage of the upper triangular unipotent matrices under
$$G(\mathcal{O}_{F_v^+}) \to G(k_v) \underset{\iota_w}{\overset{\sim}{\to}} \mathrm{GL}_2(k_v),$$
where w is a place of F over v.
- v does not split completely in $F(\zeta_p)$; that is, $(\mathbf{N}v) \not\equiv 1 \pmod{p}$.
- The ratio of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_v)$ is not equal to $(\mathbf{N}v)^{\pm 1}$.

Finally, we assume that U_{v_1} is not maximal for some place $v_1 \nmid p$ of F^+ such that the following holds.

- For any nontrivial root of unity ζ in a quadratic extension of F, v_1 does not divide $\zeta + \zeta^{-1} - 2$.

Note that, under these assumptions, U is sufficiently small. Note also that, by [**DDT97**, Lemma 4.11], it is always possible to choose a place v_1 which satisfies these hypotheses.

4.1.2. Continue to let E be a sufficiently large finite extension of $\mathbb{Q}_p$ with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$, and assume in particular that E is large enough that $\bar{r}$ is defined over $\mathbb{F}$. As in Section 3, we have a fixed set of places $\widetilde{S}_p$ of F dividing p, and we let $\widetilde{I}_p$ denote the set of embeddings $F \hookrightarrow E$ giving rise to an element of $\widetilde{S}_p$. Let R denote the set of places v of F^+ for which $U_v \neq G(\mathcal{O}_{F_v^+})$, write $T = S_p \coprod R$, and define the Hecke $\mathcal{O}$-algebra $\mathbb{T}^{T,\mathrm{univ}}$ as above.

Fix a weight $\lambda \in (\mathbb{Z}^2_+)^{\widetilde{I}_p}$, and, for each place $v \in S_p$, fix an inertial type τ_v of $I_{F_{\tilde{v}}}$.

We note that our assumptions on $\bar{r}$ and on U_v at the places v at which U_v is not maximal imply that $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ is either 0 or is locally free over $\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}[1/p]$ of rank $2^{|R|}$ (see [**Tay06**, Lemma 1.6(2)]; the requisite multiplicity one result is given by [**Lab11**, Theorems 5.4 and 5.9]).

4.2. Deformations to $\mathcal{G}_2$.

Let $\mathcal{G}_2$ be as in Section 3.2.3, and extend $\bar{r}$ to a representation $\overline{\rho}: G_{F^+}\to\mathcal{G}_2(\mathbb{F})$, with $\nu\circ\overline{\rho}=\overline{\varepsilon}^{-1}$ and $\overline{\rho}|_{G_F}=(\bar{r},\overline{\varepsilon}^{-1})$, as in Section 3.2.3.

4.2.1. Let $\mathcal{C}_{\mathcal{O}}$ denote the category of complete local Noetherian $\mathcal{O}$-algebras with residue field isomorphic to $\mathbb{F}$ via the structure map. Let S be a set of places of F^+ which split in F, containing all places dividing p. Regard $\overline{\rho}$ as a representation of $G_{F^+,S}$. As in [**CHT08**, Definition 1.2.1], we define the following.

- A *lifting* of $\overline{\rho}$ to an object A of $\mathcal{C}_{\mathcal{O}}$ is a continuous homomorphism $\rho: G_{F^+,S}\to\mathcal{G}_2(A)$ lifting $\overline{\rho}$ and with $\nu\circ\rho=\varepsilon^{-1}$.
- Two liftings ρ, ρ' of $\overline{\rho}$ to A are *equivalent* if they are conjugate by an element of $\ker(\mathrm{GL}_2(A)\to\mathrm{GL}_2(\mathbb{F}))$.
- A *deformation* of $\overline{\rho}$ to an object A of $\mathcal{C}_{\mathcal{O}}$ is an equivalence class of liftings.

Similarly, if $T\subset S$, we define the following.

- A *T-framed lifting* of $\overline{\rho}$ to A is a tuple $(\rho,\{\alpha_v\}_{v\in T})$, where ρ is a lifting of $\overline{\rho}$ and $\alpha_v\in\ker(\mathrm{GL}_2(A)\to\mathrm{GL}_2(\mathbb{F}))$ for $v\in T$.
- Two T-framed liftings $(\rho,\{\alpha_v\}_{v\in T})$, $(\rho',\{\alpha'_v\}_{v\in T})$ are *equivalent* if there is an element $\beta\in\ker(\mathrm{GL}_2(A)\to\mathrm{GL}_2(\mathbb{F}))$ with $\rho'=\beta\rho\beta^{-1}$ and $\alpha'_v=\beta\alpha_v$ for $v\in T$.
- A *T-framed deformation* of $\overline{\rho}$ is an equivalence class of T-framed liftings.

4.2.2. For each place $v\in T$, we choose a place $\widetilde{v}$ of F above v, extending the choices made for $v\in S_p$. Let $\widetilde{T}$ denote the set of places $\widetilde{v}$, $v\in T$. For each $v\in T$, we let $R^{\square}_{\widetilde{v}}$ denote the universal $\mathcal{O}$-lifting ring of $\bar{r}|_{G_{F_{\widetilde{v}}}}$. For each $v\in S_p$, write $R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}$ for $R^{\square,\lambda_v,\tau_v}_{\bar{r}|_{G_{F_{\widetilde{v}}}}}$.

We now recall from [**CHT08**, Sections 2.2 and 2.3] the notion of a *deformation problem*

$$\mathcal{S}':=(L/L^+,T',\widetilde{T}',\mathcal{O},\bar{r},\chi,\{R_{\mathcal{S}',\widetilde{v}}\}_{v\in T'}).$$

This data consists of

- an imaginary CM field L with maximal totally real subfield L^+;
- a finite set of finite places T' of L^+, each of which splits in L;
- a finite set of finite places $\widetilde{T}'$ of L, consisting of exactly one place lying over each place in T';
- the ring of integers $\mathcal{O}$ of a finite extension E of $\mathbb{Q}_p$ (assumed sufficiently large);
- $\bar{r} : G_{L^+,T'} \to \mathcal{G}_2(\mathbb{F})$ a continuous homomorphism such that $\bar{r}^{-1}(\mathrm{GL}_2(\mathbb{F}) \times \mathrm{GL}_1(\mathbb{F})) = G_{L,T'}$, and $\bar{r}|_{G_{L,T'}}$ is irreducible;
- $\chi : G_{L^+,T'} \to \mathcal{O}^\times$ a continuous character lifting $\nu \circ \bar{r}$;
- for each place $v \in T'$, a quotient $R_{\mathcal{S}',\tilde{v}}$ of $R_{\tilde{v}}^{\square}$ by an $1 + M_2(\mathfrak{m}_{R_{\tilde{v}}^{\square}})$-invariant ideal.

For any deformation problem $\mathcal{S}'$ as above, there is a universal deformation $\mathcal{O}$-algebra $R_{\mathcal{S}'}^{\mathrm{univ}}$ and a universal deformation $r_{\mathcal{S}'}^{\mathrm{univ}} : G_{L^+,T'} \to \mathcal{G}_2(R_{\mathcal{S}'}^{\mathrm{univ}})$ of $\bar{r}$, which is universal for deformations r of $\bar{r}$ with $\nu \circ r = \chi$ which satisfy the additional property that, for each $v \in T'$, the point of $\operatorname{Spec} R_{\tilde{v}}^{\square}$ corresponding to $r|_{G_{F_{\tilde{v}}}}$ is a point of $\operatorname{Spec} R_{\mathcal{S}',\tilde{v}}$. The $1 + M_2(\mathfrak{m}_{R_{\tilde{v}}^{\square}})$-invariance of $R_{\mathcal{S}',\tilde{v}}$ implies that this condition does not depend on the choice of r in its equivalence class. For any $T \subset T'$, we also consider the universal T-framed lifting ring $R_{\mathcal{S}'}^{\square_T}$, which is universal for liftings of type $\mathcal{S}'$ together with choices of basis at the places in T (see [CHT08, Definitions 2.2.1 and 2.2.7]).

Returning to our specific situation, consider the deformation problem

$$\mathcal{S} := (F/F^+, T, \widetilde{T}, \mathcal{O}, \overline{\rho}, \varepsilon^{-1}, \{R_{\tilde{v}}^{\square}\}_{v \in R} \cup \{R_{\tilde{v}}^{\square,\lambda_v,\tau_v}\}_{v \in S_p}).$$

(The quotients $R_{\tilde{v}}^{\square,\lambda_v,\tau_v}$ satisfy the condition above by [GG12, Lemma 3.2.3].) There is a corresponding universal deformation $\rho_{\mathcal{S}}^{\mathrm{univ}} : G_{F^+,T} \to \mathcal{G}_2(R_{\mathcal{S}}^{\mathrm{univ}})$ of $\overline{\rho}$.

The lifting of Theorem 3.2.4 and the universal property of $\rho_{\mathcal{S}}^{\mathrm{univ}}$ gives an $\mathcal{O}$-homomorphism

$$R_{\mathcal{S}}^{\mathrm{univ}} \twoheadrightarrow \mathbb{T}_{\lambda,\tau}^{T}(U, \mathcal{O})_{\mathfrak{m}},$$

which is surjective by Theorem 3.2.4(3).

4.3. Patching

4.3.1. In order to apply the Taylor–Wiles–Kisin method, and in particular to choose the auxiliary primes used in the patching argument, it is necessary to make an assumption on the image of the global mod p representation $\bar{r}$. In our setting, it will be convenient for us to use the notion of an *adequate* subgroup of $\mathrm{GL}_2(\overline{\mathbb{F}}_p)$, which is defined in [Tho12]. We will not need to make use of the actual definition; instead, we recall the following classification.

PROPOSITION 4.3.2. *Suppose that $p > 2$ is a prime, and that $G \subset \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a finite subgroup which acts irreducibly on $\overline{\mathbb{F}}_p^2$. Then precisely one of the following is true.*

- *We have $p = 3$, and the image of G in $\mathrm{PGL}_2(\overline{\mathbb{F}}_3)$ is conjugate to $\mathrm{PSL}_2(\mathbb{F}_3)$.*
- *We have $p = 5$, and the image of G in $\mathrm{PGL}_2(\overline{\mathbb{F}}_5)$ is conjugate to $\mathrm{PSL}_2(\mathbb{F}_5)$.*
- *G is adequate.*

Proof. This is [BLGG13b, Proposition A.2.1]. □

4.3.3. Assume from now on that the following holds.

- $\bar{r}(G_{F(\zeta_p)})$ is adequate.

We wish to consider auxiliary sets of primes in order to apply the Taylor–Wiles–Kisin patching method. Let $(Q, \widetilde{Q}, \{\overline{\psi}_{\tilde{v}}\}_{v \in Q})$ be a triple where the following hold.

- Q is a finite set of finite places of F^+ which is disjoint from T and consists of places which split in F.
- $\widetilde{Q}$ consists of a single place $\tilde{v}$ of F above each place v of F^+.
- For each $v \in Q$, $\bar{r}|_{G_{F_{\tilde{v}}}} \cong \overline{\psi}_{\tilde{v}} \oplus \overline{\psi}'_{\tilde{v}}$, where $\overline{\psi}_{\tilde{v}} \neq \overline{\psi}'_{\tilde{v}}$, and $\mathbf{N}v \equiv 1 \pmod{p}$.

For each $v \in Q$, let $R_v^{\overline{\psi}_{\tilde{v}}}$ denote the quotient of $R_v^{\square}$ corresponding to lifts $r : G_{F_{\tilde{v}}} \to \mathrm{GL}_2(A)$ which are $\ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(\mathbb{F}))$-conjugate to a lift of the form $\psi \oplus \psi'$, where ψ is a lift of $\overline{\psi}_{\tilde{v}}$ and ψ' is an unramified lift of $\overline{\psi}'_{\tilde{v}}$. We let $\mathcal{S}_Q$ denote the deformation problem

$$\mathcal{S}_Q = (F/F^+, T \cup Q, \widetilde{T} \cup \widetilde{Q}, \mathcal{O}, \overline{\rho}, \varepsilon^{-1}, \{R_{\tilde{v}}^{\square}\}_{v \in R} \cup \{R_{\tilde{v}}^{\square, \lambda_v, \tau_v}\}_{v \in S_p} \cup \{R^{\overline{\psi}_{\tilde{v}}}\}_{v \in Q}).$$

We let $R_{\mathcal{S}_Q}^{\mathrm{univ}}$ denote the corresponding universal deformation ring, and we let $R_{\mathcal{S}_Q}^{\square_T}$ denote the corresponding universal T-framed deformation ring.

We define

$$R^{\mathrm{loc}} := \left(\widehat{\otimes}_{v \in S_p} R_{\tilde{v}}^{\square, \lambda_v, \tau_v}\right) \widehat{\otimes} \left(\widehat{\otimes}_{v \in R} R_{\tilde{v}}^{\square}\right),$$

where all completed tensor products are taken over $\mathcal{O}$.

REMARK 4.3.4. Let $v \in R$. Since we have assumed that the ratio of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_v)$ is not equal to $(\mathbf{N}v)^{\pm 1}$, and $(\mathbf{N}v) \not\equiv 1 \pmod p$, $R^{\square}_{\bar{r}|_{G_{L_{\tilde{v}}}}}$ is formally smooth of relative dimension four over $\mathcal{O}$ (this may be checked by computing the dimension of the reduced tangent space by the usual Galois cohomology calculation; see [CHT08, Lemma 2.4.9]), and in particular all deformations of $\bar{r}|_{G_{L_{\tilde{v}}}}$ are unramified. Applying Proposition 2.1.1 above and [BLGHT11, Lemma 3.3], we see that R^{loc} is equidimensional of dimension $1 + 4|T| + [F^+ : \mathbb{Q}]$, and $R^{\mathrm{loc}}[1/p]$ is formally smooth.

4.3.5. For each finite place w of F, let $U_0(w)$ be the subgroup of $\mathrm{GL}_2(\mathcal{O}_{F,w})$ consisting of matrices congruent to $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$ modulo w, and let $U_1(w)$ be the subgroup of $\mathrm{GL}_2(\mathcal{O}_{F,w})$ consisting of matrices congruent to $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$ modulo w. For $i = 0, 1$, let $U_i(Q) = \prod_v U_i(Q)_v$ be the compact open subgroups of $G(\mathbb{A}^{\infty}_{F^+})$ defined by $U_i(Q)_v = U_v$ if $v \notin Q$, and $U_i(Q)_v = \iota_{\tilde{v}}^{-1} U_i(\tilde{v})$ if $v \in Q$. We have natural maps

$$\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_1(Q), \mathcal{O}) \twoheadrightarrow \mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_0(Q), \mathcal{O}) \twoheadrightarrow \mathbb{T}^{T \cup Q}_{\lambda,\tau}(U, \mathcal{O}) \hookrightarrow \mathbb{T}^{T}_{\lambda,\tau}(U, \mathcal{O}).$$

Note that $\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U, \mathcal{O})_{\mathfrak{m}} = \mathbb{T}^{T}_{\lambda,\tau}(U, \mathcal{O})_{\mathfrak{m}}$ by [CHT08, Proof of Corollary 3.4.5]. By taking its image, $\mathfrak{m}$ determines maximal ideals of the first three algebras in this sequence which we denote by $\mathfrak{m}_Q$ for the first two and $\mathfrak{m}$ for the third.

4.3.6. Let $\rho_{\mathfrak{m}_Q} : G_{F^+, T \cup Q} \to \mathrm{GL}_2(\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_1(Q), \mathcal{O})_{\mathfrak{m}_Q})$ be the representation defined in Theorem 3.2.4. For each $v \in Q$, choose an element $\varphi_{\tilde{v}} \in G_{F_{\tilde{v}}}$ lifting the geometric Frobenius element of $G_{F_{\tilde{v}}}/I_{F_{\tilde{v}}}$, and let $\varpi_{\tilde{v}} \in \mathcal{O}_{F_{\tilde{v}}}$ be the uniformizer with $\mathrm{Art}_{F_{\tilde{v}}} \varpi_{\tilde{v}} = \varphi_{\tilde{v}}|_{F^{\mathrm{ab}}_{\tilde{v}}}$. Let $P_{\tilde{v}}(X) \in \mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_1(Q), \mathcal{O})_{\mathfrak{m}_Q}[X]$ denote the characteristic polynomial of $\rho_{\mathfrak{m}_Q}(\varphi_{\tilde{v}})$. Since $\overline{\psi}_{\tilde{v}}(\varphi_{\tilde{v}}) \neq \overline{\psi}'_{\tilde{v}}(\varphi_{\tilde{v}})$, by Hensel's lemma we can factor $P_{\tilde{v}}(X) = (X - A_{\tilde{v}})(X - B_{\tilde{v}})$, where $A_{\tilde{v}}, B_{\tilde{v}} \in \mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_1(Q), \mathcal{O})_{\mathfrak{m}_Q}$ lift $\overline{\psi}_{\tilde{v}}(\varphi_{\tilde{v}})$ and $\overline{\psi}'_{\tilde{v}}(\varphi_{\tilde{v}})$, respectively.

For $i = 0, 1$ and $\alpha \in F^{\times}_{\tilde{v}}$ of nonnegative valuation, consider the Hecke operator

$$V_\alpha := \iota_{\tilde{v}}^{-1}\left(\left[U_i(\tilde{v}) \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} U_i(\tilde{v})\right]\right)$$

on $S_{\lambda,\tau}(U_i(Q), \mathcal{O})$. Denote by $\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_i(Q), \mathcal{O})' \subset \mathrm{End}_{\mathcal{O}}(S_{\lambda,\tau}(U_i(Q), \mathcal{O}))$ the $\mathcal{O}$-subalgebra generated by $\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_i(Q), \mathcal{O})$ and the $V_{\varpi_{\tilde{v}}}$ for $v \in Q$. We denote by $\mathfrak{m}'_Q$ the maximal ideal of $\mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_i(Q), \mathcal{O})'$ generated by $\mathfrak{m}_Q$ and the $V_{\varpi_{\tilde{v}}} - A_{\tilde{v}}$. Write $\mathbb{T}_{i,Q} := \mathbb{T}^{T \cup Q}_{\lambda,\tau}(U_i(Q), \mathcal{O})'_{\mathfrak{m}'_Q}$.

Let Δ_Q denote the maximal p-power order quotient of $U_0(Q)/U_1(Q)$. Let $\mathfrak{a}_Q$ denote the kernel of the augmentation map $\mathcal{O}[\Delta_Q] \to \mathcal{O}$. Exactly as in the proof of the sublemma [BLGG11, Theorem 3.6.1], we have the following.

(1) The natural map

$$\prod_{v\in Q}(V_{\varpi_{\tilde{v}}} - B_{\tilde{v}}) : S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \to S_{\lambda,\tau}(U_0(Q),\mathcal{O})_{\mathfrak{m}'_Q}$$

is an isomorphism.

(2) $S_{\lambda,\tau}(U_1(Q),\mathcal{O})_{\mathfrak{m}'_Q}$ is free over $\mathcal{O}[\Delta_Q]$ with

$$S_{\lambda,\tau}(U_1(Q),\mathcal{O})_{\mathfrak{m}'_Q}/\mathfrak{a}_Q \xrightarrow{\sim} S_{\lambda,\tau}(U_0(Q),\mathcal{O})_{\mathfrak{m}'_Q} \xrightarrow{\sim} S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}.$$

(3) For each $v \in Q$, there is a character with open kernel $V_{\tilde{v}} : F_{\tilde{v}}^{\times} \to \mathbb{T}_{1,Q}^{\times}$ so that the following hold.

(a) For each element $\alpha \in F_{\tilde{v}}^{\times}$ of nonnegative valuation, $V_{\alpha} = V_{\tilde{v}}(\alpha)$ on $S_{\lambda,\tau}(U_1(Q),\mathcal{O})_{\mathfrak{m}'_Q}$.

(b) $(\rho_{\mathfrak{m}_Q} \otimes_{\mathbb{T}^{T\cup Q}_{\lambda,\tau}(U_1(Q),\mathcal{O})_{\mathfrak{m}_Q}} \mathbb{T}_{1,Q})|_{W_{F_{\tilde{v}}}} \cong \psi' \oplus (V_{\tilde{v}} \circ \mathrm{Art}_{F_{\tilde{v}}}^{-1})$ with ψ' an unramified lift of $\overline{\psi}'_{\tilde{v}}$ and $(V_{\tilde{v}} \circ \mathrm{Art}_{F_{\tilde{v}}}^{-1})$ lifting $\overline{\psi}_{\tilde{v}}$.

4.3.7. The above shows, in particular, that the lift $\rho_{\mathfrak{m}_Q} \otimes \mathbb{T}_{1,Q}$ of $\bar{\rho}$ is of type $\mathcal{S}_Q$ and gives rise to a surjection $R^{\mathrm{univ}}_{\mathcal{S}_Q} \twoheadrightarrow \mathbb{T}_{1,Q}$. We think of $S_{\lambda,\tau}(U_1(Q),\mathcal{O})_{\mathfrak{m}_Q}$ as an $R^{\mathrm{univ}}_{\mathcal{S}_Q}$-module via this map.

Thinking of Δ_Q as the image of the product of the inertia subgroups in the maximal abelian p-power order quotient of $\prod_{v\in Q} G_{F_{\tilde{v}}}$, the determinant of any choice of universal deformation $r^{\mathrm{univ}}_{\mathcal{S}_Q}$ gives rise to a homomorphism $\Delta_Q \to (R^{\mathrm{univ}}_{\mathcal{S}_Q})^{\times}$. We thus have homomorphisms

$$\mathcal{O}[\Delta_Q] \to R^{\mathrm{univ}}_{\mathcal{S}_Q} \to R^{\square_T}_{\mathcal{S}_Q}$$

and natural isomorphisms $R^{\mathrm{univ}}_{\mathcal{S}_Q}/\mathfrak{a}_Q \cong R^{\mathrm{univ}}_{\mathcal{S}}$ and $R^{\square_T}_{\mathcal{S}_Q}/\mathfrak{a}_Q \cong R^{\square_T}_{\mathcal{S}}$. (This follows from (3)(b) above, which shows that, for each place $v \in Q$, the ramification of $r^{\mathrm{univ}}_{\mathcal{S}_Q}$ at $\tilde{v}$ is given by the character $(V_{\tilde{v}} \circ \mathrm{Art}_{F_{\tilde{v}}}^{-1})$.)

4.3.8. We have assumed that $\bar{r}(G_{F(\zeta_p)})$ is adequate, or equivalently (because we are considering two-dimensional representations) big in the terminology of [**CHT08**]. By [**CHT08**, Proposition 2.5.9], this implies that we can (and do) choose an integer $q \geqslant [F^+ : \mathbb{Q}]$ and for each $N \geqslant 1$ a tuple $(Q_N, \widetilde{Q}_N, \{\overline{\psi}_{\tilde{v}}\}_{v\in Q_N})$ as above such that the following hold.

- $\#Q_N = q$ for all N.
- $\mathbf{N}v \equiv 1 (\text{mod } p^N)$ for $v \in Q_N$.
- The ring $R^{\square_T}_{\mathcal{S}_{Q_N}}$ can be topologically generated over R^{loc} by $q - [F^+ : \mathbb{Q}]$ elements.

We will apply the above constructions to each of these tuples $(Q_N, \widetilde{Q}_N, \{\overline{\psi}_{\widetilde{v}}\}_{v\in Q_N})$.

Choose a lift $r^{\text{univ}}_{\mathcal{S}} : G_{F^+,S} \to \mathcal{G}_2(R^{\text{univ}}_{\mathcal{S}})$ representing the universal deformation. Let

$$\mathcal{T} = \mathcal{O}[[X_{v,i,j} : v \in T, i, j = 1, 2]].$$

The tuple $(r^{\text{univ}}_{\mathcal{S}}, (1_2 + X_{v,i,j})_{v\in T})$ (where $1_2 + X_{v,i,j}$ is a 2×2 matrix) gives rise to an isomorphism $R^{\square_T}_{\mathcal{S}} \xrightarrow{\sim} R^{\text{univ}}_{\mathcal{S}} \widehat{\otimes}_{\mathcal{O}} \mathcal{T}$. (Note that the action of j in the group $\mathcal{G}_2$ implies that this tuple has no nontrivial scalar endomorphisms.) For each N, choose a lift $r^{\text{univ}}_{\mathcal{S}_{Q_N}} : G_{F^+} \to \mathcal{G}_2(R^{\text{univ}}_{\mathcal{S}_{Q_N}})$ representing the universal deformation, with $r^{\text{univ}}_{\mathcal{S}_{Q_N}} \bmod \mathfrak{a}_{Q_N} = r^{\text{univ}}_{\mathcal{S}}$. This gives rise to an isomorphism $R^{\square_T}_{\mathcal{S}_{Q_N}} \xrightarrow{\sim} R^{\text{univ}}_{\mathcal{S}_{Q_N}} \widehat{\otimes}_{\mathcal{O}} \mathcal{T}$ which reduces modulo $\mathfrak{a}_{Q_N}$ to the isomorphism $R^{\square_T}_{\mathcal{S}} \xrightarrow{\sim} R^{\text{univ}}_{\mathcal{S}} \widehat{\otimes}_{\mathcal{O}} \mathcal{T}$.

We let

$$\begin{aligned} M &= S_{\lambda,\tau}(U, \mathcal{O})_{\mathfrak{m}} \\ M_N &= S_{\lambda,\tau}(U_1(Q_N), \mathcal{O})_{\mathfrak{m}'_{Q_N}} \otimes_{R^{\text{univ}}_{\mathcal{S}_{Q_N}}} R^{\square_T}_{\mathcal{S}_{Q_N}}. \end{aligned}$$

Then M_N is a finite free $\mathcal{T}[\Delta_{Q_N}]$-module with $M_N/\mathfrak{a}_{Q_N} \cong M \otimes_{R^{\text{univ}}_{\mathcal{S}}} R^{\square_T}_{\mathcal{S}}$, compatibly with the isomorphism $R^{\square_T}_{\mathcal{S}_{Q_N}}/\mathfrak{a}_{Q_N} \cong R^{\square_T}_{\mathcal{S}}$.

Fix a filtration by $\mathbb{F}$-subspaces

$$0 = L_0 \subset L_1 \subset \cdots \subset L_s = L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F}$$

such that each L_i is $G(\mathcal{O}_{F^+_p})$-stable, and, for each $i = 0, 1, \ldots, s-1$, $\sigma_i := L_{i+1}/L_i$ is an absolutely irreducible representation of $G(\mathcal{O}_{F^+_p})$. This in turn induces a filtration on $S_{\lambda,\tau}(U, \mathcal{O})_{\mathfrak{m}} \otimes_{\mathcal{O}} \mathbb{F}$ (respectively $S_{\lambda,\tau}(U_1(Q_N), \mathcal{O})_{\mathfrak{m}_{Q_N}} \otimes_{\mathcal{O}} \mathbb{F}$) whose graded pieces are the finite-dimensional $\mathbb{F}$-vector spaces $S(U, \sigma_i)_{\mathfrak{m}}$ (respectively the finite free $\mathbb{F}[\Delta_{Q_N}]$-modules $S(U_1(Q_N), \sigma_i)_{\mathfrak{m}_{Q_N}}$). By extension of scalars, we obtain a filtration on $M_N \otimes_{\mathcal{O}} \mathbb{F}$. We denote these filtrations by

$$0 = M^0 \subset M^1 \subset \cdots \subset M^s = M \otimes_{\mathcal{O}} \mathbb{F}$$

and

$$0 = M^0_N \subset M^1_N \subset \cdots \subset M^s_N = M_N \otimes_{\mathcal{O}} \mathbb{F}.$$

Let $g = q - [F^+ : \mathbb{Q}]$, let

$$\begin{aligned} \Delta_\infty &= \mathbb{Z}_p^q, \\ R_\infty &= R^{\text{loc}}[[x_1, \ldots, x_g]], \\ R'_\infty &= \left(\widehat{\otimes}_{v \in T} R^{\square}_{\tilde{v}}\right)[[x_1, \ldots, x_g]], \\ S_\infty &= \mathcal{T}[[\Delta_\infty]], \end{aligned}$$

and let $\mathfrak{a}$ denote the kernel of the $\mathcal{O}$-algebra homomorphism $S_\infty \to \mathcal{O}$ which sends each $X_{v,i,j}$ to 0 and each element of Δ_∞ to 1. Note that S_∞ is formally smooth over $\mathcal{O}$ of relative dimension $q + 4|T|$, and that R_∞ is a quotient of R'_∞. For each N, choose a surjection $\Delta_\infty \twoheadrightarrow \Delta_{Q_N}$, and let $\mathfrak{c}_N$ denote the kernel of the corresponding homomorphism $S_\infty \twoheadrightarrow \mathcal{T}[\Delta_{Q_N}]$. For each $N \geqslant 1$, choose a surjection of R^{loc}-algebras

$$R_\infty \twoheadrightarrow R^{\square_T}_{\mathcal{S}_{Q_N}}.$$

We regard each $R^{\square_T}_{\mathcal{S}_{Q_N}}$ as an S_∞-algebra via $S_\infty \twoheadrightarrow \mathcal{T}[\Delta_{Q_N}] \to R^{\square_T}_{\mathcal{S}_{Q_N}}$. In particular, $R^{\square_T}_{\mathcal{S}_{Q_N}}/\mathfrak{a} \cong R^{\text{univ}}_{\mathcal{S}}$.

Now a patching argument as in [Kis09a, 2.2.9] shows that there exist the following:

- an $\mathcal{O}$-module homomorphism $S_\infty \to R_\infty$, and an R_∞-module M_∞ which is finite free as an S_∞-module;
- a filtration by R_∞-modules
$$0 = M^0_\infty \subset M^1_\infty \subset \cdots \subset M^s_\infty = M_\infty \otimes_{\mathcal{O}} \mathbb{F}$$
whose graded pieces are finite free $S_\infty/\pi S_\infty$-modules;
- a surjection of R^{loc}-algebras $R_\infty/\mathfrak{a}R_\infty \to R^{\text{univ}}_{\mathcal{S}}$; and
- an isomorphism of R_∞-modules $M_\infty/\mathfrak{a}M_\infty \xrightarrow{\sim} M$ which identifies M^i with $M^i_\infty/\mathfrak{a}M^i_\infty$.

We claim that we can make the above construction so that, for $i = 1, 2, \ldots, s$, the (R'_∞, S_∞)-bimodule $M^i_\infty/M^{i-1}_\infty$ and the isomorphism $M^i_\infty/(\mathfrak{a}M^i_\infty + M^{i-1}_\infty) \xrightarrow{\sim} M^i/M^{i-1}$ depend only on $(U, \mathfrak{m}$ and) the isomorphism class of L_i/L_{i-1} as a $G(\mathcal{O}_{F_p^+})$-representation, but not on (λ, τ). For any finite collection of pairs (λ, τ), this follows by the same finiteness argument used during patching. Since the set of (λ, τ) is countable, the claim follows from a diagonalization argument.

For σ a global Serre weight, we denote by M^σ_∞ the $R_\infty/\pi R_\infty$-module constructed above when $L_i/L_{i-1} \xrightarrow{\sim} \sigma$, and we set

$$\mu'_\sigma(\bar{r}) = 2^{-|R|} e_{R_\infty/\pi}(M^\sigma_\infty).$$

LEMMA 4.3.9. *For each σ, $\mu'_\sigma(\bar{r})$ is a nonnegative integer. Moreover, the following conditions are equivalent.*

(1) *The support of M meets every irreducible component of* $\operatorname{Spec} R^{\mathrm{loc}}[1/p]$.

(2) *$M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a faithfully flat $R_\infty[1/p]$-module which is locally free of rank $2^{|R|}$.*

(3) *R_S^{univ} is a finite $\mathcal{O}$-algebra and $M \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a faithful $R_S^{\mathrm{univ}}[1/p]$-module.*

(4)
$$e(R_\infty/\pi R_\infty) = \sum_{i=1}^{s} 2^{-|R|} e_{R_\infty/\pi}(M_\infty^{\sigma_i}) = \sum_{i=1}^{s} \mu'_{\sigma_i}(\bar{r}),$$
where σ_i is a global Serre weight with $L_i/L_{i-1} \xrightarrow{\sim} \sigma_i$.

Proof. We argue in a similar fashion to the proof of Lemma 2.2.11 of [Kis09a]. By Remark 4.3.4, $R_\infty[1/p]$ is formally smooth of dimension $q + 4|T| = \dim S_\infty[1/p]$. Since M_∞ is free over S_∞, the module $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ has depth $q + 4|T|$ at every maximal ideal of $R_\infty[1/p]$ in its support. By the Auslander–Buchsbaum formula, $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ has projective dimension zero, and as it is finite over $S_\infty[1/p]$ it is also finite and therefore finite flat over $R_\infty[1/p]$.

If $Z \subset \operatorname{Spec} R_\infty[1/p]$ is an irreducible component in the support of M_∞, then Z is finite over $\operatorname{Spec} S_\infty[1/p]$ and of the same dimension. Hence the map $Z \to \operatorname{Spec} S_\infty[1/p]$ is surjective. As $(M_\infty/\mathfrak{a}M_\infty)[1/p] = M[1/p]$ has rank $2^{|R|}$ over any point of $R_S^{\mathrm{univ}}[1/p]$ in its support, $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ has rank $2^{|R|}$ over Z. This shows that (1) and (2) are equivalent.

Let $\operatorname{Spec} A \subset \operatorname{Spec} R_\infty$ denote the closure of the support of M_∞, on $\operatorname{Spec} R_\infty[1/p]$. By what we have just seen, there exists a map $A^{2^{|R|}} \to M_\infty$, which is an isomorphism at the generic points of $\operatorname{Spec} A$. Using [Kis09a, Proposition 1.3.4], we see that
$$2^{-|R|} e_{R_\infty/\pi}(M_\infty/\pi M_\infty) = 2^{-|R|} e_{A/\pi}(M_\infty/\pi M_\infty) = e(A/\pi)$$
is an integer. If $\sigma = \sigma_a$ is a global Serre weight, as in Section 3.3, then applying the above with $(\lambda, \tau) = (\lambda_a, 1)$ shows that $\mu'_\sigma(\bar{r})$ is an integer.

We also have
$$e(R_\infty/\pi R_\infty) \geqslant 2^{-|R|} e_{R_\infty/\pi}(M_\infty/\pi M_\infty) = \sum_{i=1}^{s} 2^{-|R|} e_{R_\infty/\pi}(M_\infty^{\sigma_i})$$
with equality if and only if M_∞ is a faithful R_∞-module or, equivalently, if and only if the support of M meets every irreducible component of $R_\infty[1/p]$. So (1) and (4) are equivalent.

Finally, if M_∞ is a faithful R_∞-module, then R_∞ is finite over S_∞, and so R_S^{univ}, which is a quotient of $R_\infty/\mathfrak{a}$, is a finite $\mathcal{O}$-module. This shows that (2) implies (3). The converse is a consequence of the Khare–Wintenberger argument; see [**KW09**, Theorem 3.3]. To be precise, assuming R_S^{univ} is a finite $\mathcal{O}$-algebra, we see that the image of

$$\mathrm{Spec}\, R_S^{\mathrm{univ}} \to \mathrm{Spec}\, R^{\mathrm{loc}}$$

meets every component of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$, because it follows from [**BLGGT14a**, Proposition 1.5.1(2)] that the quotient R_S^{univ} corresponding to any particular component of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$ is a finite $\mathcal{O}$-algebra of dimension at least one, and therefore has $\overline{\mathbb{Q}}_p$-points. Hence, if $M \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a faithful R_S^{univ}-module, then the support of M meets every component of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$, which is (1). □

Let $\sigma = \otimes_{v \in S_p} \sigma_v$ be a global Serre weight. The following lemma will be useful in order to determine the $\mu'_\sigma(\bar{r})$ more precisely in some situations.

LEMMA 4.3.10. *The multiplicity $\mu'_\sigma(\bar{r})$ is nonzero if and only if $S(U,\sigma)_\mathfrak{m} \neq 0$. If this holds, then, for each place $v \mid p$ of F, $\bar{r}|_{G_{F_v}}$ has a crystalline lift of Hodge type σ_v in the sense of Section 2.2.3.*

Proof. By definition, $\mu'_\sigma(\bar{r}) \neq 0$ if and only if $M_\infty^\sigma \neq 0$. Moreover, $M_\infty^\sigma \neq 0$ if and only if $M_\infty^\sigma/\mathfrak{a}M_\infty^\sigma \neq 0$, and $M_\infty^\sigma/\mathfrak{a}M_\infty^\sigma \cong S(U,\sigma)_\mathfrak{m}$ by definition, so we indeed have $\mu'_\sigma(\bar{r}) \neq 0$ if and only if $S(U,\sigma)_\mathfrak{m} \neq 0$.

For the second part, let $a = (a_v)_{v \in S_p}$ with $a_v \in (\mathbb{Z}_+^2)_{\mathrm{sw}}^{\mathrm{Hom}\,(k_v,\mathbb{F})}$ and $\sigma_v \cong \sigma_{a_v}$. Note that, since U is sufficiently small, we have $S(U,\sigma)_\mathfrak{m} \neq 0$ if and only if $S_{\lambda_a,1}(U,\mathcal{O}) \neq 0$, where λ_a is defined in Section 3.3, and 1 denotes the trivial type. The result then follows at once from Theorem 3.2.4(4). □

4.4. Potential diagonalizability. We now use the methods of [**BLGGT14a**] to show that the equivalent conditions of Lemma 4.3.9 are frequently achieved. We begin by recalling the definition of *potential diagonalizability*, a notion defined in [**BLGGT14a**]. We will use this definition here for convenience, as it allows us to make use of certain results from [**BLGG13b**], and to easily argue simultaneously in the potentially Barsotti–Tate and Fontaine–Laffaille cases.

Suppose that $K/\mathbb{Q}_p$ is a finite extension with residue field k, that $E/\mathbb{Q}_p$ is a finite extension with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$, and that $\rho_1, \rho_2 : G_K \to \mathrm{GL}_2(\mathcal{O})$ are two continuous representations. We say that ρ_1 *connects* to ρ_2 if all of the following hold.

- ρ_1 and ρ_2 are both crystalline of the same Hodge type λ.
- $\overline{\rho}_1 \cong \overline{\rho}_2$.
- ρ_1 and ρ_2 define points on the same irreducible component of $R^{\square,\lambda}_{\overline{\rho}_1} \otimes_{\mathcal{O}} \overline{\mathbb{Q}}_p$.

We say that $\rho : G_K \to \mathrm{GL}_2(\mathcal{O})$ is *diagonal* if it is a direct sum of crystalline characters, and we say that ρ is *diagonalizable* if it connects to some diagonal representation. Finally, we say that ρ is *potentially diagonalizable* if there is a finite extension L/K such that $\rho|_{G_L}$ is diagonalizable. We say that a representation $G_K \to \mathrm{GL}_2(E)$ is potentially diagonalizable if the representation on some G_K-invariant lattice is potentially diagonalizable; this is independent of the choice of lattice by [**BLGGT14a**, Lemma 1.4.1].

The following two lemmas, which rely on our earlier papers, are the key to our applications of Lemma 4.3.9 to the Breuil–Mézard conjecture for potentially Barsotti–Tate representations.

LEMMA 4.4.1. *If $\rho : G_K \to \mathrm{GL}_2(E)$ is potentially Barsotti–Tate, then it is potentially diagonalizable.*

Proof. Choose a finite extension L/K such that L contains a primitive pth root of unity, and $\overline{\rho}|_{G_L}$ is trivial. Then $\overline{\rho}|_{G_L}$ has a decomposable ordinary crystalline lift of Hodge type zero, namely $\rho_1 := 1 \oplus \varepsilon^{-1}$. Extending L if necessary, we may also assume that $\overline{\rho}|_{G_L}$ has a decomposable nonordinary crystalline lift of Hodge type zero, say ρ_2 (this is simply a direct sum of Lubin–Tate characters). By [**Gee06**, Proposition 2.3] and [**Kis09b**, Corollary 2.5.16] we see that $\rho|_{G_L}$ connects to one of ρ_1, ρ_2, and in either case ρ is potentially diagonalizable by definition. □

LEMMA 4.4.2. *Let $a \in (\mathbb{Z}^2_+)^{\mathrm{Hom}\,(k_v,\mathbb{F})}_{\mathrm{sw}}$ and $\sigma = \sigma_a$ the corresponding Serre weight.*

(1) *If σ is not a predicted Serre weight for $\bar{r}$, then $R^{\square,\sigma}_{\bar{r}} = 0$. In particular, it is vacuously the case that every crystalline representation $\rho : G_K \to \mathrm{GL}_2(E)$ of Hodge type a which lifts $\overline{\rho}$ is potentially diagonalizable.*

(2) *If $K/\mathbb{Q}_p$ is unramified and σ is regular in the sense of Definition 2.2.4, then every crystalline representation $\rho : G_K \to \mathrm{GL}_2(E)$ of Hodge type a which lifts $\overline{\rho}$ is potentially diagonalizable.*

Proof. In the case that σ is not a predicted Serre weight for $\overline{\rho}$, the main result of [**GLS13**] shows that there are no crystalline lifts of $\overline{\rho}$ of Hodge type a, and the result follows.

If $K/\mathbb{Q}_p$ is unramified and σ is regular, then the result follows immediately from the main theorem of [GL12]. □

We will apply these results by using the following corollary of the methods of [BLGGT14a]. In Corollary 4.4.3 below, we maintain the notation and assumptions made throughout this section.

COROLLARY 4.4.3. *Suppose that* $\bar{r} : G_F \to \mathrm{GL}_2(\mathbb{F})$ *satisfies the assumptions of Sections 4.1.1 and 4.3.3, and that, for each place* $v \mid p$*, every lift of* $\bar{r}|_{G_{F_{\tilde{v}}}}$ *of Hodge type* λ_v *and Galois type* τ_v *is potentially diagonalizable. Then the equivalent conditions of Lemma 4.3.9 hold.*

Proof. We will show that condition (1) of Lemma 4.3.9 holds, that is, that the support of M meets every irreducible component of $\operatorname{Spec} R^{\mathrm{loc}}[1/p]$. By the correspondence between our algebraic automorphic forms and automorphic forms on GL_2 (which is explained in detail in [BLGG13b, Section 2]), this is equivalent to the statement that, if we make for each place $v \mid p$ of F^+ any choice of component $\operatorname{Spec} R_{\tilde{v}}$ of $\operatorname{Spec} R_{\tilde{v}}^{\square,\lambda_v,\tau_v}[1/p]$, there is a continuous lift $r : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ of $\bar{r}$ such that the following hold.

- $r^c \equiv r^\vee \varepsilon^{-1}$.
- r is unramified at all places not dividing p (note that this will be automatic at the places in R by Lemma 4.3.4).
- For each place $v \mid p$ of F^+, $r|_{G_{F_{\tilde{v}}}}$ corresponds to a point of $R_{\tilde{v}}$.
- r is automorphic in the sense of [BLGGT14a].

By [BLGG13b, Lemma 3.1.1], we may choose a solvable extension F_1/F of CM fields such that the following hold.

- F_1 is linearly disjoint from $\bar{F}^{\ker \mathrm{ad}\,\bar{r}}(\zeta_p)$ over F.
- There is a continuous lift $r' : G_{F_1} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ of $\bar{r}|_{G_{F_1}}$ such that r' is automorphic, and, for each place $w|p$ of F_1, $r'|_{G_{F_{1,w}}}$ is (potentially) diagonalizable.

The existence of r now follows by applying [BLGG13b, Theorem A.4.1] (which is another variant of the Khare–Wintenberger argument; again, see [KW09, Theorem 3.3]) with the representation $r_{l,\iota}(\pi')$ in the statement of [KW09, Theorem 3.3] equal to r'. It is here that we use the assumption that every lift of $\bar{r}|_{G_{F_{\tilde{v}}}}$ of Hodge type λ_v and Galois type τ_v is potentially diagonalizable, as we need to know that the points of $\operatorname{Spec} R_{\tilde{v}}$ correspond to potentially diagonalizable lifts in order to satisfy the hypotheses of [**KW09**, Theorem A.4.1] (note that, since $r|_{G_{F_1}}$ is automorphic and the extension F_1/F is solvable, r is automorphic, by [**BLGHT11**, Lemma 1.4]). □

4.5. Local results. We will now combine Corollary 4.4.3 with the local-to-global results of Appendix A to prove our main local results. We begin with a lemma from linear algebra, for which we need to establish some notation. Given a vector space V over $\mathbb{Q}$ with a choice of basis, we let $V_{\geqslant 0}$ denote the cone spanned by nonnegative linear combinations of the basis elements. If V, W are vector spaces over $\mathbb{Q}$ with choices of bases, then we will choose the corresponding tensor basis for $V \otimes W$, and define $(V \otimes W)_{\geqslant 0}$ accordingly. For any set I, we set $\mathbb{Z}^I_{\geqslant 0} = \mathbb{Z}^I \cap \mathbb{Q}^I_{\geqslant 0}$. In particular, $(\mathbb{Z}^m)^{\otimes n}_{\geqslant 0} = (\mathbb{Z}^m)^{\otimes n} \cap (\mathbb{Q}^m)^{\otimes n}_{\geqslant 0}$.

LEMMA 4.5.1. *Let k be a field.*

(1) *If for $i = 1, \ldots, n$ we have injective linear maps $\alpha_i : V_i \hookrightarrow W_i$ between k-vector spaces, then $\alpha_1 \otimes \cdots \otimes \alpha_n : V_1 \otimes \cdots \otimes V_n \to W_1 \otimes \cdots \otimes W_n$ is also injective.*

(2) *If for $i = 1, \ldots, n$ we have linear maps $\alpha_i : V_i \to W_i$ between k-vector spaces and nonzero elements $w_i \in W_i$ such that*

$$w_1 \otimes \cdots \otimes w_n \in \operatorname{Im}(\alpha_1 \otimes \cdots \otimes \alpha_n),$$

then, for each i, $w_i \in \operatorname{Im}(\alpha_i)$.

(3) *Let I be a (possibly infinite) set, and let $\alpha : \mathbb{Z}^m_{\geqslant 0} \to \mathbb{Z}^I_{\geqslant 0}$ be a map that extends to an injective linear map $\alpha : \mathbb{Q}^m \to \mathbb{Q}^I$. Suppose that $v \in \mathbb{Q}^m$, with $\alpha(v) \in \mathbb{Z}^I_{\geqslant 0}$, and that, for some $n \geqslant 1$, v satisfies $v^{\otimes n} \in (\mathbb{Z}^m)^{\otimes n}_{\geqslant 0}$. Then $v \in \mathbb{Z}^m_{\geqslant 0}$.*

Proof. (1) By induction, it suffices to treat the case $n = 2$. Then we can factor $\alpha_1 \otimes \alpha_2$ as the composite

$$V_1 \otimes V_2 \to W_1 \otimes V_2 \to W_1 \otimes W_2$$

of two maps which are each injective.

(2) For each $j \neq i$, choose $\varphi_j \in W_j^*$ with $\varphi_j(w_j) = 1$. Identifying V_j and W_j with $k \otimes \cdots \otimes k \otimes V_j \otimes \cdots \otimes k$ and $k \otimes \cdots \otimes k \otimes W_j \otimes \cdots \otimes k$, respectively, we see that, if

$$w_1 \otimes \cdots \otimes w_n = (\alpha_1 \otimes \cdots \otimes \alpha_n)(\vec{v}),$$

then

$$w_i = \alpha_i((\varphi_1\alpha_1 \otimes \cdots \otimes \varphi_{i-1}\alpha_{i-1} \otimes 1 \otimes \varphi_{i+1}\alpha_{i+1} \otimes \cdots \otimes \varphi_n\alpha_n)(\vec{v})),$$

as required.

(3) By explicitly examining the entries of $v^{\otimes n}$ in the standard basis, one sees easily that $v^{\otimes n} \in (\mathbb{Z}^m)^{\otimes n}_{\geqslant 0}$ implies that either v or $-v$ is in $\mathbb{Z}^m_{\geqslant 0}$. If $-v \in \mathbb{Z}^m_{\geqslant 0}$, then $\alpha(v), -\alpha(v) \in \mathbb{Z}^I_{\geqslant 0}$, which implies that $\alpha(v) = 0$, and $v = 0$ as α is injective. $\square$

In order to apply this result, we will make use of the following lemma. The last assertion (allowing the determinant of τ to run over tame characters) will be used in Section 5.

LEMMA 4.5.2. *Let p be a prime, let $K/\mathbb{Q}_p$ be a finite extension with residue field k, and let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. Then the system of equations*

$$e(R^{\square,0,\tau}_{\bar{r}}/\pi) = \sum_\sigma n_{0,\tau}(\sigma)\mu_\sigma(\bar{r}),$$

in the unknowns $\mu_\sigma(\bar{r})$ has at most one solution. Equivalently, the linear map which sends $(\mu_\sigma)_\sigma$ to $(\sum_\sigma n_{0,\tau}(\sigma)\mu_\sigma)_{0,\tau}$ is injective. In fact, this is true even if we restrict consideration to the set of types τ for which $\det\tau$ *is tame.*

Proof. If L is a topological field, and G is a topological group, let $R_L(G)$ denote the Grothendieck group of continuous L-representations of G.

The composite of reduction mod p and semisimplification yields a homomorphism $\pi : R_E(\mathrm{GL}_2(\mathcal{O}_K)) \to R_{\mathbb{F}}(\mathrm{GL}_2(k))$. Since by definition this homomorphism takes $\sigma(\tau) = W_0 \otimes \sigma(\tau)$ to $\sum n_{0,\tau}(\sigma)\sigma$, it is enough to check that the $\pi(\sigma(\tau))$ span $R_{\mathbb{F}}(\mathrm{GL}_2(k))$. The surjection $\mathrm{GL}_2(\mathcal{O}_K) \twoheadrightarrow \mathrm{GL}_2(k)$ gives an injection $R_E(\mathrm{GL}_2(k)) \hookrightarrow R_E(\mathrm{GL}_2(\mathcal{O}_K))$, and, by [Ser77, Theorem 33 of Ch. 16], the homomorphism $R_E(\mathrm{GL}_2(k)) \to R_{\mathbb{F}}(\mathrm{GL}_2(k))$ is surjective; so π is certainly surjective.

We recall the explicit classification of irreducible E-representations of $\mathrm{GL}_2(k)$; see [Dia07, Section 1]. There are the one-dimensional representations $\chi \circ \det$, the twists St_χ of the Steinberg representation, the principal series representations $I(\chi_1, \chi_2)$, and the cuspidal representations $\Theta(\xi)$. By the explicit construction in Henniart's appendix to [BM02], all but the representations St_χ occur as a $\sigma(\tau)$ for some tame type τ (the principal series representations occur for tamely ramified types of niveau one, and the cuspidal types for tamely ramified types of niveau two), so, in order to complete the proof, it is enough to check that the $\pi(\mathrm{St}_\chi)$ are in the span of the $\pi(\sigma(\tau))$, where τ runs over the representations with tame determinant.

To see this, note that the reduction mod p of St_χ is just the irreducible representation $\overline{\chi} \circ \det \otimes \sigma_{p-1,\ldots,p-1}$ of $\mathrm{GL}_2(k)$. Let $\psi : \mathcal{O}_K^\times \to E^\times$ be a nonquadratic ramified character with trivial reduction, let $\tilde{\omega}$ denote the Teichmüller lift of ω, and consider the element

$$\sigma_\chi := \sigma(\chi\psi \oplus \chi\psi^{-1}) - \sigma(\chi\psi\tilde{\omega} \oplus \chi\psi^{-1}\tilde{\omega}^{-1}) + \sigma(\chi\tilde{\omega} \oplus \chi\tilde{\omega}^{-1}) - \sigma(\chi \oplus \chi)$$

of $R_E(\mathrm{GL}_2(\mathcal{O}_K))$. By [**BD13**, Proposition 4.2], $\pi(\sigma_{\chi^{-1}}) = \overline{\chi} \circ \det \otimes \sigma_{p-1,\ldots,p-1}$, as required. □

COROLLARY 4.5.3. *Suppose a solution to the equations in Lemma 4.5.2 exists.*

- *If* $\sigma|_{k^\times} \neq (\overline{\varepsilon} \det \bar{r})^{-1} \circ \mathrm{Art}_K$, *then* $\mu_\sigma(\bar{r}) = 0$.
- *The* $\mu_\sigma(\bar{r})$ *for which* $\sigma|_{k^\times} = (\overline{\varepsilon} \det \bar{r})^{-1} \circ \mathrm{Art}_K$ *are determined by the equations corresponding to the types* τ *with determinant* $\widetilde{\overline{\varepsilon} \det \bar{r}}$.

Proof. By Lemma 4.5.2, we certainly need only consider the equations with tame determinant. Now, if $\det \tau$ does not lift $\overline{\varepsilon} \det \bar{r}$, then certainly $R_{\bar{r}}^{\square,0,\tau} = 0$. It is also easy to check that $n_{0,\tau}(\sigma) = 0$ unless the central character of σ is $(\det \overline{\tau})^{-1} \circ \mathrm{Art}_K$, so that, if we set $\mu_\sigma(\bar{r}) = 0$ when the central character of σ is not equal to $(\overline{\varepsilon} \det \bar{r})^{-1} \circ \mathrm{Art}_K$, then all the equations for types τ with $\det \overline{\tau} \neq \overline{\varepsilon} \det \bar{r}$ will automatically be satisfied.

Furthermore, none of the equations with $\det \tau = \widetilde{\overline{\varepsilon} \det \bar{r}}$ involve any of these values of $\mu_\sigma(\bar{r})$, so, since the equations have a unique solution by Lemma 4.5.2, it follows that we must have $\mu_\sigma(\bar{r}) = 0$ if the central character of σ is not equal to $(\overline{\varepsilon} \det \bar{r})^{-1} \circ \mathrm{Art}_K$, and that the remaining values of $\mu_\sigma(\bar{r})$ are determined by the τ with $\det \tau = \widetilde{\overline{\varepsilon} \det \bar{r}}$. Note that, if $p > 2$, then it follows from twisting that, if the equations hold for all τ with $\det \tau = \widetilde{\overline{\varepsilon} \det \bar{r}}$, then in fact they hold for all τ with $\det \overline{\tau} = \overline{\varepsilon} \det \bar{r}$. □

REMARK 4.5.4. A referee has pointed out that in fact the obvious variant of Lemma 4.5.2 where the types and weights have fixed central character may be proved via [**BD13**, Propositions 4.2 and 4.3] (or, in the case when $p > 2$, it can be deduced directly from the previous remark by twisting).

We may now combine Corollary 4.4.3 with Lemmas 4.5.1 and 4.5.2 and Corollary A.5 to prove our main local result for potentially diagonalizable representations.

THEOREM 4.5.5. *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be a finite extension with residue field k, and let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. Then there are uniquely determined nonnegative integers $\mu_\sigma(\bar{r})$, σ running over local Serre weights, with the following property: for any pair (λ, τ) with the property that every potentially crystalline lift of $\bar{r}$ of Hodge type λ and Galois type τ is potentially diagonalizable,*

$$e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi) = \sum_{\sigma} n_{\lambda,\tau}(\sigma)\mu_\sigma(\bar{r}).$$

Proof. By Corollaries A.5 and 4.4.3, we see that all of the equivalent conditions of Lemma 4.3.9 hold in a case where each $F_{\tilde{v}} \cong K$ for each prime $v \mid p$ of F, and each $\bar{r}|_{G_{F_{\tilde{v}}}}$ is an unramified twist of a representation isomorphic to our local $\bar{r}$. In particular, condition (4) of Lemma 4.3.9 holds. In this case, by Lemma 2.1.2 we see that

$$e(R_\infty/\pi R_\infty) = \prod_{v|p} e(R_{\bar{r}}^{\square,\lambda_v,\tau_v}/\pi) = \sum_{\{\sigma_v\}_{v|p}} \left(\prod_{v|p} n_{\lambda_v,\tau_v}(\sigma_v)\right)\mu'_{\sigma_{\mathrm{gl}}}(\bar{r}),$$

where in the final expression the sum runs over tuples $\{\sigma_v\}_{v|p}$ of equivalence classes of Serre weights, and $\sigma_{\mathrm{gl}} \cong \otimes_{v|p}\sigma_v$.

Choose an ordering of the set of equivalence classes of (local) Serre weights, and denote its cardinality by m. Let I be the set of pairs (λ, τ) in the statement of the theorem. We will apply Lemma 4.5.1 to the map $\alpha : \mathbb{Q}^m \to \mathbb{Q}^I$ which sends $(\mu_\sigma)_\sigma$ to $(\sum_\sigma n_{\lambda,\tau}(\sigma)\mu_\sigma)_{\lambda,\tau}$. Note that all the pairs $(0, \tau)$ are in I, by Lemma 4.4.1, so that α is an injective map by Lemma 4.5.2, and induces a map $\mathbb{Z}^m_{\geqslant 0} \to \mathbb{Z}^I_{\geqslant 0}$ since the $n_{\lambda,\tau}(\sigma)$ are nonnegative integers.

Let $w = (e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi))_{\lambda,\tau} \in \mathbb{Z}^I_{\geqslant 0}$, and take n to be the number of primes of F^+ lying over p. By what we saw above, there exists $v_n \in (\mathbb{Z}^m)^{\otimes n}_{\geqslant 0}$ such that $\alpha^{\otimes n}(v_n) = w^{\otimes n}$. Hence $w = \alpha(v)$ for some $v \in \mathbb{Q}^m$, by Lemma 4.5.1(2). As $\alpha^{\otimes n}$ is injective, Lemma 4.5.1(1) implies that we must have $v_n = v^{\otimes n}$, so that $v \in \mathbb{Z}^m_{\geqslant 0}$, by Lemma 4.5.1(3). Defining the $\mu_\sigma(\bar{r})$ by $v = (\mu_\sigma(\bar{r}))_\sigma$, the theorem follows. □

COROLLARY 4.5.6 (Theorem A). *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be a finite extension with residue field k, and let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. Then there are uniquely determined nonnegative integers $\mu_\sigma(\bar{r})$ such that, for all inertial types τ, we have*

$$e(R_{\bar{r}}^{\square,0,\tau}/\pi) = \sum_{\sigma} n_{0,\tau}(\sigma)\mu_\sigma(\bar{r}).$$

Furthermore, the $\mu_\sigma(\bar{r})$ enjoy the following properties.

(1) *$\mu_\sigma(\bar{r}) \neq 0$ if and only if σ is a predicted Serre weight for $\bar{r}$.*

(2) *If $K/\mathbb{Q}_p$ is unramified and σ is regular, then $\mu_\sigma(\bar{r}) = e(R_{\bar{r}}^{\square,\sigma}/\pi)$, where $R_{\bar{r}}^{\square,\sigma}$ is the crystalline lifting ring of Hodge type determined by σ. If furthermore σ is Fontaine–Laffaille regular, then $\mu_\sigma(\bar{r}) = 1$ if σ is a predicted Serre weight for $\bar{r}$, and is zero otherwise.*

Proof. The initial part is an immediate consequence of Theorem 4.5.5 and Lemma 4.4.1. The numbered properties also follow easily from the results above, as we now show. Recall that, in the proof of Theorem 4.5.5, we worked with a global representation $\bar{r}$ with the property that, for each $v \mid p$, $\bar{r}|_{G_{F_{\tilde{v}}}}$ was an unramified twist of our local $\bar{r}$. We will continue to use this global representation $\bar{r}$. Let σ be an equivalence class of local Serre weights, and let a be a global Serre weight such that $\sigma_a \cong \sigma \otimes \cdots \otimes \sigma$. By the proof of Theorem 4.5.5, we see that $\mu'_\sigma(\bar{r}) = \mu_\sigma(\bar{r})^n$, where there are n places of F^+ lying over p. In particular, we have $\mu'_\sigma(\bar{r}) \neq 0$ if and only if $\mu_\sigma(\bar{r}) \neq 0$.

(1) Lemma 4.3.10 shows that $\mu_\sigma(\bar{r}) \neq 0$ if and only if (the global representation) $\bar{r}$ is modular of weight $\sigma \otimes \cdots \otimes \sigma$. The main result of [**BLGG13b**] shows that, whenever σ is a predicted Serre weight for $\bar{r}$, then $\bar{r}$ is modular of weight $\sigma \otimes \cdots \otimes \sigma$. The converse holds by the main result of [**GLS13**].

(2) That $\mu_\sigma(\bar{r}) = e(R_{\bar{r}}^{\square,\sigma}/\pi)$ whenever σ is regular is an immediate consequence of Theorem 4.5.5 and Lemma 4.4.2(2). Now suppose that σ is Fontaine–Laffaille regular. By Fontaine–Laffaille theory (or as a special case of the main result of [**GLS14**]), $R_{\bar{r}}^{\square,\sigma} \neq 0$ if and only if σ is a predicted Serre weight for $\bar{r}$, so to complete the proof it is enough to show that, if σ is Fontaine–Laffaille regular and $R_{\bar{r}}^{\square,\sigma} \neq 0$, then $e(R_{\bar{r}}^{\square,\sigma}/\pi) = 1$. By [**CHT08**, Lemma 2.4.1] (see also [**CHT08**, Definition 2.2.6]) $R_{\bar{r}}^{\square,\sigma}$ is formally smooth over $\mathcal{O}$, so the result follows. □

REMARK 4.5.7. If (as seems plausible) it is always the case that any crystalline lift of $\bar{r}$ of Hodge type determined by σ is in fact potentially diagonalizable, then it would follow that $\mu_\sigma(\bar{r}) = e(R_{\bar{r}}^{\square,\sigma}/\pi)$ (without any assumption on K or σ).

In the next section, it will be helpful to have the following definition.

DEFINITION 4.5.8. Let $W^{\mathrm{BT}}(\bar{r})$ be the set of Serre weights σ for which $\mu_\sigma(\bar{r}) > 0$.

Recall from Section 2 that sets of Serre weights $W^{\text{explicit}}(\bar{r})$ and $W^{\text{cris}}(\bar{r})$ are defined in [**BLGG13b**], and that $W^{\text{cris}}(\bar{r})$ is simply the set of Serre weights σ for which $\bar{r}$ has a crystalline lift of Hodge type σ. (Recall that in the ramified case this notion depends on the choice of particular embeddings $K \hookrightarrow E$; the following corollary is true for any such choice.)

COROLLARY 4.5.9. *We have equalities* $W^{\text{explicit}}(\bar{r}) = W^{\text{BT}}(\bar{r}) = W^{\text{cris}}(\bar{r})$.

Proof. The equality $W^{\text{explicit}}(\bar{r}) = W^{\text{BT}}(\bar{r})$ is an immediate consequence of Corollary 4.5.6(1) and the definitions of $W^{\text{explicit}}(\bar{r})$ and $W^{\text{BT}}(\bar{r})$. By the main result of [**GLS13**], we have $W^{\text{explicit}}(\bar{r}) = W^{\text{cris}}(\bar{r})$, as required. □

5. The Buzzard–Diamond–Jarvis Conjecture

5.1. Types. We now apply the machinery developed above to the weight part of Serre's conjecture for inner forms of GL_2 (as opposed to the outer forms treated in [**BLGG13b**, **GLS12**, **GLS14**, **GLS13**]).

We briefly recall some results from Henniart's appendix to [**BM02**] which will be useful to us in what follows. Let $l \neq p$ be prime, and let L be a finite extension of $\mathbb{Q}_l$ with ring of integers $\mathcal{O}_L$ and residue field k_L. Let $\tau : I_L \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be an inertial type. Then Henniart defines an irreducible finite-dimensional representation $\sigma(\tau)$ of $\mathrm{GL}_2(\mathcal{O}_L)$ with the following property.

- If π is an infinite-dimensional smooth irreducible $\overline{\mathbb{Q}}_p$-representation of $\mathrm{GL}_2(L)$, then $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_L)}(\sigma(\tau), \pi^\vee) \neq 0$ if and only if $r_p(\pi^\vee)|_{I_L} \cong \tau$, in which case $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_L)}(\sigma(\tau), \pi^\vee)$ is one dimensional.

(Note that, in contrast to the case $l = p$ covered in Theorem 2.1.3, we make no prescription on the monodromy. The only difference between the two definitions is for scalar inertial types, where we replace a twist of the trivial representation with a twist of the small Steinberg representation.)

For later use, we need to understand the basic properties of the reductions modulo p of the $\sigma(\tau)$. We would like to thank Guy Henniart for his assistance with the proof of the following lemma.

LEMMA 5.1.1. *Let $\overline{\pi}$ be an infinite-dimensional, admissible smooth irreducible $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_2(L)$. Then there is an inertial type τ such that for any $\mathrm{GL}_2(\mathcal{O}_L)$-stable $\bar{\mathbb{Z}}_p$-lattice $L_\tau \subset \sigma(\tau)$ we have $(L_\tau \otimes_{\bar{\mathbb{Z}}_p} \overline{\pi})^{\mathrm{GL}_2(\mathcal{O}_L)} \neq 0$.*

Proof. By [**Vig89b**, Corollaire 13], $\overline{\pi}$ may be lifted to an admissible smooth, irreducible $\overline{\mathbb{Q}}_p$-representation π of $\mathrm{GL}_2(L)$. Set $\tau = r_p(\pi^\vee)|_{I_L}$. Then we have

$\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_L)}(\sigma(\tau), \pi^\vee) \neq 0$, and hence

$$(\sigma(\tau) \otimes \pi)^{\mathrm{GL}_2(\mathcal{O}_L)} = \mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_L)}(\pi^\vee, \sigma(\tau)) \neq 0$$

as the category of finite-dimensional smooth $\bar{\mathbb{Q}}_p$-representations of $\mathrm{GL}_2(\mathcal{O}_L)$ is semisimple. The lemma follows. □

5.1.2. We will also need the analogue of this result for (nonsplit) quaternion algebras. Let D be the quaternion algebra with centre L, and let π be a smooth irreducible (so finite-dimensional) $\overline{\mathbb{Q}}_p$-representation of $D^\times$. Let $\mathcal{O}_D$ be the maximal order in D, so that $L^\times\mathcal{O}_D^\times$ has index two in $D^\times$. Thus $\pi|_{\mathcal{O}_D^\times}$ is either irreducible or a sum of two distinct irreducible representations which are conjugate under a uniformizer in $D^\times$, and we easily see that if π' is another smooth irreducible representation of $D^\times$, then π and π' differ by an unramified twist if and only if $\pi|_{\mathcal{O}_D^\times} \cong \pi'|_{\mathcal{O}_D^\times}$.

Let τ be as above, and assume that τ is either irreducible or scalar. Then there is an irreducible smooth representation π_τ of $D^\times$ such that $r_p(\mathrm{JL}(\pi_\tau))|_{I_L} \cong \tau$, where JL denotes the Jacquet–Langlands correspondence, and any two such representations differ by an unramified twist (see Section A.1.3 of Henniart's appendix to [**BM02**]). Define $\sigma_D(\tau)$ to be an irreducible constituent of $\pi_\tau^\vee|_{\mathcal{O}_D^\times}$; then, by the above discussion, we have the following property.

- If π is a smooth irreducible $\overline{\mathbb{Q}}_p$-representation of $D^\times$, then $\mathrm{Hom}_{\mathcal{O}_D^\times}(\sigma_D(\tau), \pi^\vee)$ is nonzero if and only if $r_p(\mathrm{JL}(\pi))|_{I_L} \cong \tau$, in which case $\mathrm{Hom}_{\mathcal{O}_D^\times}(\sigma_D(\tau), \pi^\vee)$ is one dimensional.

We also have the following analogue of Lemma 5.1.1.

Lemma 5.1.3. *Let $\overline{\pi}$ be an admissible smooth irreducible $\overline{\mathbb{F}}_p$-representation of $D^\times$. Then there an inertial type τ such that for any $\mathcal{O}_D^\times$-stable $\bar{\mathbb{Z}}_p$-lattice $L_\tau \subset \sigma_D(\tau)$ we have $(L_\tau \otimes \overline{\pi})^{\mathcal{O}_D^\times} \neq 0$.*

Proof. By [**Vig89a**, Théorème 4], $\overline{\pi}$ may be lifted to an admissible smooth irreducible $\overline{\mathbb{Q}}_p$-representation π of $D^\times$, and the result follows as in the proof of Lemma 5.1.1. □

5.2. Deformation rings. Assume from now on that $p > 2$. We now carry out our global patching argument. Since the arguments are by now rather standard, and in any case extremely similar to those of Section 4, we will sketch the construction, giving the necessary definitions and explaining the differences from the arguments of Section 4. For a detailed treatment of the Taylor–Wiles–Kisin method for Shimura curves, the reader could consult [**BD13**].

5.2.1. For technical reasons, we will fix the determinants of all the deformations we will consider, and we now introduce some notation to allow this. We will also need to consider p-adic representations of the absolute Galois groups of l-adic fields. Accordingly, let $E/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$, let l be a prime (possibly equal to p), let $K/\mathbb{Q}_l$ be a finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. We will always assume that E is sufficiently large that all representations under consideration are defined over E. Fix a finite order character $\psi : G_K \to E^\times$ such that $\det \bar{r} = \bar{\varepsilon}^{-1}\overline{\psi}$. Let $\tau : I_K \to \mathrm{GL}_2(E)$ be an inertial type such that $\det \tau = \psi|_{I_K}$. Recall that we have the universal $\mathcal{O}$-lifting ring $R_{\bar{r}}^{\square}$ of $\bar{r}$, and let $R_{\bar{r}}^{\square,\psi}$ denote the universal $\mathcal{O}$-lifting ring for liftings of determinant $\psi\varepsilon^{-1}$.

Suppose first that $l = p$. Let $R_{\bar{r}}^{\square,\lambda,\tau,\psi}$ be the unique quotient of $R_{\bar{r}}^{\square,\lambda,\tau}$ with the property that the $\overline{\mathbb{Q}}_p$-points of $R_{\bar{r}}^{\square,\lambda,\tau,\psi}$ are precisely the $\overline{\mathbb{Q}}_p$-points of $R_{\bar{r}}^{\square,\lambda,\tau}$ whose associated Galois representations have determinant $\psi\varepsilon^{-1}$.

Suppose now that $l \neq p$. Then, by [Gee11a, Theorem 2.1.6], there is a unique (possibly zero) p-torsion free reduced quotient $R_{\bar{r}}^{\square,\tau,\psi}$ of $R_{\bar{r}}^{\square}$ whose $\overline{\mathbb{Q}}_p$-points are precisely those points of $R_{\bar{r}}^{\square}$ which correspond to liftings of $\bar{r}$ which have determinant $\psi\varepsilon^{-1}$ and Galois type τ, in the sense that the restriction to I_K of the corresponding Weil–Deligne representations are isomorphic to τ.

REMARK 5.2.2. By [EG14, Lemma 4.3.1], we have an isomorphism

$$R_{\bar{r}}^{\square,\lambda,\tau} \cong R_{\bar{r}}^{\square,\lambda,\tau,\psi}[[X]],$$

so that in particular we have $e(R_{\bar{r}}^{\square,\lambda,\tau,\psi}/\pi) = e(R_{\bar{r}}^{\square,\lambda,\tau}/\pi)$.

5.2.3. We begin with a generalization of [Gee11a, Corollary 3.1.7] (see also [BD13, Théorème 3.2.2]), which produces modular liftings of mod p Galois representations with prescribed local properties, and will be used in place of Corollary 4.4.3 in this setting. The proof is very similar to that of Corollary 4.4.3. We first define some notation.

Let F be a totally real field, and let $\overline{\rho} : G_F \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation. We will say that $\overline{\rho}$ is *modular* if it is isomorphic to the reduction mod p of the Galois representation associated to a Hilbert modular eigenform of parallel weight two. Fix a totally even, finite order character $\psi : G_F \to E^\times$ with the property that $\det \overline{\rho} = \overline{\psi}\bar{\varepsilon}^{-1}$. Let S be a finite set of finite places of F, including all places dividing p and all places at which $\overline{\rho}$ or ψ is ramified. For each place $v \in S$, fix an inertial type τ_v of I_{F_v} such that $\det \tau_v = \psi|_{I_{F_v}}$. For each place $v \mid p$, let R_v be a quotient of $R_{\overline{\rho}|_{G_{F_v}}}^{\square,0,\tau_v,\psi|_{G_{F_v}}}$ corresponding to a choice of an irreducible component of Spec $R^{\square,0,\tau_v,\psi|_{G_{F_v}}}_{\overline{\rho}|_{G_{F_v}}}[1/p]$, and, for each place $v \in S$ with $v \nmid p$, let R_v be a quotient of $R^{\square,\tau_v,\psi|_{G_{F_v}}}_{\overline{\rho}|_{G_{F_v}}}$ corresponding to a choice of an irreducible component of Spec $R^{\square,\tau_v,\psi|_{G_{F_v}}}_{\overline{\rho}|_{G_{F_v}}}[1/p]$.

5.2.4. We assume from now on that $\bar{\rho}$ satisfies the following conditions.

- $\overline{\rho}$ is modular.
- $\overline{\rho}|_{G_{F(\zeta_p)}}$ is absolutely irreducible.
- If $p = 5$, the projective image of $\overline{\rho}|_{G_{F(\zeta_p)}}$ is not isomorphic to A_5.

LEMMA 5.2.5. *For $\bar{\rho}$ and R_v as above, there is a continuous lift $\rho : G_F \to \mathrm{GL}_2(E)$ of $\overline{\rho}$ such that the following hold.*

- $\det \rho = \psi\varepsilon^{-1}$.
- *ρ is unramified outside of S.*
- *For each place $v \in S$, $\rho|_{G_{F_v}}$ arises from a point of $R_v[1/p]$.*
- *ρ is modular.*

Proof. This is essentially an immediate consequence of [**Gee11a**, Corollary 3.1.7], given the main result of [**BLGG13a**]. Indeed, in the case that for all places $v \mid p$ the component R_v corresponds to nonordinary representations, the lemma is a special case of [**Gee11a**, Corollary 3.1.7], and the only thing to be checked in general is that the hypothesis (ord) of [**Gee11a**, Proposition 3.1.5] is satisfied. This hypothesis relates to the existence of ordinary lifts of $\overline{\rho}$, and a general result on the existence of such lifts is proved in [**BLGG13a**].

In fact, examining [**Gee11a**, Proof of Proposition 3.1.5], we see that it is enough to check that there is some finite solvable extension L/F of totally real fields with the property that L is linearly disjoint from $\overline{F}^{\ker \overline{\rho}}(\zeta_p)$ over F, and $\overline{\rho}|_{G_L}$ has a modular lift which is potentially Barsotti–Tate and ordinary at all places $v \mid p$. In order to see that such an L exists, simply choose a solvable totally real extension L/F such that L is linearly disjoint from $\overline{F}^{\ker \overline{\rho}}(\zeta_p)$ over F and $\overline{\rho}|_{G_{L_w}}$ is reducible for each place $w|p$ of L. The existence of the required lift of $\overline{\rho}|_{G_L}$ is then immediate from the main result of [**BLGG13a**]. ☐

5.2.6. Now, let Σ, $\Sigma' \subset S$ be disjoint subsets, not containing any places dividing p. Suppose that if $v \in \Sigma$ then τ_v is either irreducible or scalar, and if $v \in \Sigma'$ then τ_v is scalar.

For each $v \in S$, $v \nmid p$, we define $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}}$ as follows: $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}} = R^{\square,\tau_v,\psi}_{\overline{\rho}|G_{F_v}}$ unless $v \in \Sigma \cup \Sigma'$ and τ_v is scalar. If τ_v is scalar, and $v \in \Sigma$ (respectively, $v \in \Sigma'$) then we define $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}}$ as the quotient of $R^{\square,\tau_v,\psi}_{\overline{\rho}|G_{F_v}}$ corresponding to the components of $R^{\square,\tau_v,\psi}_{\overline{\rho}|G_{F_v}}[1/p]$ whose $\overline{\mathbb{Q}}_p$-points are not all potentially unramified (respectively, are all potentially unramified). Note that at this stage we could have $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}} = 0$. For example, if $v \in \Sigma$ and τ_v is scalar and $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}} \neq 0$, then $\overline{\rho}|_{G_{F_v}}$ is necessarily a twist of an extension of the trivial character by the mod p cyclotomic character.

By [Pil08, Theorems 4.1.1 and 4.1.3], Spec $R^{\square,\tau_v,\psi}_{\overline{\rho}|G_{F_v}}[1/p]$ is the union of formally smooth irreducible components, and two such components C_1, C_2 can have a nontrivial intersection only in the following situation: τ_v is scalar, and there is a character $\gamma : G_{F_v} \to E^\times$ such that (up to exchanging C_1, C_2) the representations parameterized by $x \in C_1(\bar{\mathbb{Q}}_p)$ are unramified after twisting by γ^{-1}, and the representations parameterized by $x \in C_2(\bar{\mathbb{Q}}_p)$ are extensions of γ by $\gamma(1)$, which are ramified if $x \in C_2 \backslash C_1(\bar{\mathbb{Q}}_p)$.

We set $R^{\psi}_S = \widehat{\otimes}_{v\in S,\mathcal{O}} R^{\square,\psi}_{\overline{\rho}|G_{F_v}}$, $R^{\tau,\psi}_p = \widehat{\otimes}_{v|p,\mathcal{O}} R^{\square,0,\tau_v,\psi}_{\overline{\rho}|G_{F_v}}$, $R^{\tau,\psi}_{S\backslash p} = \widehat{\otimes}_{v\nmid p, v\in S,\mathcal{O}} R^{\square,\tau_v,\psi,*}_{\overline{\rho}|G_{F_v}}$, and $R^{\tau,\psi}_S = R^{\tau,\psi}_p \widehat{\otimes}_{\mathcal{O}} R^{\tau,\psi}_{S\backslash p}$. (Note that $R^{\tau,\psi}_S$ is analogous to the ring R^{loc} defined in Section 4.3.3.)

5.3. Modular forms. We continue to use the notation and assumptions introduced in the last subsection.

5.3.1. Let D be quaternion algebra with centre F. We assume that either D is ramified at all infinite places (the definite case) or that D is split at precisely one infinite place (the indefinite case), and that the set of finite primes at which D is ramified is precisely the set Σ introduced above. If $F = \mathbb{Q}$, assume further that $D \neq M_2(\mathbb{Q})$. (We make this assumption only in order to give a uniform definition of our spaces of modular forms; with the usual modifications to handle the noncompactness of modular curves, the case $D = M_2(\mathbb{Q})$ could also be treated by our methods. Since our main results are all well known in this case, we do not comment further on this assumption below.)

Fix a maximal order $\mathcal{O}_D$ of D, and an isomorphism $(\mathcal{O}_D)_v \xrightarrow{\sim} M_2(\mathcal{O}_{F_v})$ at each finite place $v \notin \Sigma$. For each finite place v of F, we let $\pi_v \in F_v$ denote a uniformizer.

By a *global Serre weight for* $D^\times$ we mean an absolutely irreducible mod p representation σ of $\prod_{v|p}(\mathcal{O}_D)^\times_v$ considered up to equivalence. Using the isomorphisms $(\mathcal{O}_D)_v \xrightarrow{\sim} M_2(\mathcal{O}_{F_v})$ for $v|p$, each such σ has the form $\sigma \cong \otimes_{v|p}\sigma_v$, where σ_v is a local Serre weight for $\mathrm{GL}_2(k_v)$, k_v the residue field of v. For the remainder of this section, when we say 'global Serre weight' we mean a global Serre weight for $D^\times$.

5.3.2. We now define the spaces of modular forms with which we will work, following Section 3 of [BD13], to which we refer for a more detailed treatment of the facts that we need, and for further references to the literature.

By [DDT97, Lemma 4.11] we can and do choose a finite place $v_1 \notin S$ of F such that ρ is unramified at v_1, $\mathbf{N}v_1 \not\equiv 1 \pmod p$, and the ratio of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_{v_1})$ is not equal to $(\mathbf{N}v_1)^{\pm 1}$. In particular, we have $H^0(G_{F_{v_1}}, \operatorname{ad}\overline{\rho}(1)) = (0)$, and it follows that any lifting of $\overline{\rho}|_{G_{F_{v_1}}}$ is necessarily unramified. Furthermore, we can and do assume that the residue characteristic of v_1 is sufficiently large that, for any nontrivial root of unity ζ in a quadratic extension of F, v_1 does not divide $\zeta + \zeta^{-1} - 2$.

Fix from now on the compact open subgroup $U = \prod_v U_v \subset (D \otimes_{\mathbb{Q}} \mathbb{A}^\infty)^\times$, where $U_v = (\mathcal{O}_D)_v^\times$ for $v \neq v_1$, and U_{v_1} is the subgroup of $\mathrm{GL}_2(\mathcal{O}_{F,v_1})$ consisting of elements which are upper triangular unipotent modulo v_1. As in the case of unitary groups, the assumption on v_1 is used (implicitly) below to guarantee that M_∞ is a free S_∞-module (see, for example, condition (2.1.2) in [Kis09a]).

For each place $v \in S$, we have an inertial type τ_v, and thus a finite-dimensional irreducible E-representation $\sigma(\tau_v)$ of $(\mathcal{O}_D)_v^\times$. (When $v \mid p$ the representation $\sigma(\tau_v)$ is defined by Theorem 2.1.3, and when $v \nmid p$ it is defined in Section 5.1, where it was denoted $\sigma_{D_v}(\tau_v)$ in the case that D_v is ramified.)

Now define a representation of $(\mathcal{O}_D)_v^\times$ on a finite free $\mathcal{O}$-module L_{τ_v}, as follows: if $v \notin \Sigma'$, then we choose L_{τ_v} to be a $(\mathcal{O}_D)_v^\times$-stable $\mathcal{O}$-lattice in $\sigma(\tau_v)$. If $v \in \Sigma'$, then τ_v is scalar, by assumption, and we set $L_{\tau_v} = \mathcal{O}$, equipped with an action of $(\mathcal{O}_D)_v^\times$ given by $\tau_v^{-1} \circ \mathrm{Art}_{F_v} \circ \det$. We write $L_{\tau^p} = \otimes_{v\in S, v\nmid p, \mathcal{O}} L_{\tau_v}$, $L_{\tau_p} = \otimes_{v|p,\mathcal{O}} L_{\tau_v}$.

5.3.3. Let θ denote a finite $\mathcal{O}$-module with a continuous action of $\prod_{v|p} U_v$, with the property that the action of $\prod_{v|p} \mathcal{O}_{F_v}^\times \subset \prod_{v|p} U_v$ on θ is given by $\psi \circ \mathrm{Art}_F$. Then $\theta \otimes_{\mathcal{O}} L_{\tau^p}$ has an action of U via the projection onto $\prod_{v\in S} U_v$. We extend this action to an action of $U(\mathbb{A}_F^\infty)^\times$ by letting $(\mathbb{A}_F^\infty)^\times$ act via the composition of the projection $(\mathbb{A}_F^\infty)^\times \to (\mathbb{A}_F^\infty)^\times / F^\times$ and $\psi \circ \mathrm{Art}_F$. (This action is well defined by our assumptions on ψ and θ.)

Let $V \subset U$ be a compact open subgroup, and suppose first that we are in the indefinite case. Then there is a smooth projective algebraic curve X_V over F associated to V, and a local system $\mathcal{F}_{\theta\otimes_{\mathcal{O}} L_{\tau^p}}$ on X_V corresponding to $\theta \otimes_{\mathcal{O}} L_{\tau^p}$, and we set

$$S_{\tau^p}(V, \theta) := H^1(X_{U,\overline{\mathbb{Q}}}, \mathcal{F}_{\theta\otimes_{\mathcal{O}} L_{\tau^p}}).$$

If we are in the definite case, then we let $S_{\tau^p}(V,\theta)$ be the set of continuous functions

$$f: D^\times\backslash(D\otimes_F\mathbb{A}_F^\infty)^\times \to \theta\otimes_{\mathcal{O}} L_{\tau^p}$$

such that we have $f(gu)=u^{-1}f(g)$ for all $g\in(D\otimes_F\mathbb{A}_F)^\times$, $u\in V(\mathbb{A}_F^\infty)^\times$.

In the special case that $\theta=L_{\tau_p}$, we write $S_\tau(V,\mathcal{O})$ for $S_{\tau^p}(V,L_{\tau_p})$.

For the precise relationship between these spaces and automorphic forms on $D^\times$, see [**Kis09b**, Section 3.1.14] (in the definite case) and [**BD13**, Section 3.5] (in the indefinite case).

5.3.4. We now define the Hecke algebras that we will use. Let T be a finite set of finite places of F containing S, and such that $(\mathcal{O}_D)_v^\times\subset V$ for $v\notin T$. Let $\mathbb{T}^{T,\mathrm{univ}}$ be the commutative $\mathcal{O}$-polynomial algebra generated by formal variables T_v, S_v for each finite place $v\notin T\cup\{v_1\}$ of F. Then $\mathbb{T}^{T,\mathrm{univ}}$ acts on $S_{\tau^p}(V,\theta)$ via the Hecke operators

$$T_v=\left[\mathrm{GL}_2(\mathcal{O}_{F,v})\begin{pmatrix}\pi_v & 0\\ 0 & 1\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_{F,v})\right],$$

$$S_v=\left[\mathrm{GL}_2(\mathcal{O}_{F,v})\begin{pmatrix}\pi_v & 0\\ 0 & \pi_v\end{pmatrix}\mathrm{GL}_2(\mathcal{O}_{F,v})\right].$$

We denote the image of $\mathbb{T}^{T,\mathrm{univ}}$ in $\mathrm{End}_{\mathcal{O}}(S_\tau(V,\mathcal{O}))$ by $\mathbb{T}_\tau^T(V,\mathcal{O})$.

Let $\mathfrak{m}$ be the maximal ideal of $\mathbb{T}^{S,\mathrm{univ}}$ with residue field $\mathbb{F}$ with the property that, for each finite place $v\notin S\cup\{v_1\}$ of F, the characteristic polynomial of $\overline{\rho}(\mathrm{Frob}_v)$ is equal to the image of $X^2-T_vX+(\mathbf{N}v)S_v$ in $\mathbb{F}[X]$. Note that, if $S_\tau(U,\mathcal{O})_{\mathfrak{m}}\neq 0$, then $S_\tau(U,\mathcal{O})_{\mathfrak{m}}\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ is locally free of rank 2^m over $\mathbb{T}_\tau^S(U,\mathcal{O})_{\mathfrak{m}}[1/p]$, where $m=1$ in the definite case and $m=2$ in the indefinite case. (This follows from our assumptions on v_1 and U, and the multiplicity one property of the $\sigma(\tau_v)$ explained in Theorem 2.1.3 and Section 5.1.)

5.3.5. Let $R_{F,S}^\psi$ be the universal deformation ring for deformations of $\bar{\rho}$ with determinant $\psi\varepsilon^{-1}$ which are unramified outside S. Then $S_\tau(U,\mathcal{O})_{\mathfrak{m}}$ is naturally a $R_{F,S}^\psi$-module. In particular, a Hecke eigenform in $S_\tau(U,\mathcal{O})_{\mathfrak{m}}$ gives rise to a deformation $\rho: G_{F,S}\to\mathrm{GL}_2(\mathcal{O})$ of $\bar{\rho}$, with determinant $\psi\varepsilon^{-1}$.

LEMMA 5.3.6. *For each $v\in S$ and $v|p$ (respectively, $v\nmid p$), let R_v be a quotient of $R_{\overline{\rho}|_{G_{F_v}}}^{\square,0,\tau_v,\psi}$ corresponding to an irreducible component of* Spec $R_{\overline{\rho}|_{G_{F_v}}}^{\square,0,\tau_v,\psi}[1/p]$ *(respectively,* Spec $R_{\overline{\rho}|_{G_{F_v}}}^{\square,\tau_v,\psi,*}$*). (In particular, we are assuming that the set of such components is nonempty for each $v\in S$.) Then there is a continuous lift $\rho: G_F\to\mathrm{GL}_2(\mathcal{O})$ of $\overline{\rho}$ such that the following hold.*

- $\det \rho = \psi\varepsilon^{-1}$.
- *ρ is unramified outside of S.*
- *For each place $v \in S$, $\rho|_{G_{F_v}}$ arises from a point of $R_v[1/p]$.*
- *ρ arises from a Hecke eigenform in $S_\tau(U, \mathcal{O})$.*

Proof. By Lemma 5.2.5 (note that we are already assuming that the hypotheses of that lemma are satisfied), there exists a ρ with the first three properties, which is modular in the sense that it arises from a Hilbert modular form π. The conditions on ρ at $v \mid p$ imply that this form is parallel of weight two.

We now check that π is discrete series at all $v \in \Sigma$, so that π transfers to an automorphic form π^D on D. Suppose that π_v is not discrete series for some $v \in \Sigma$. Since we are assuming that τ_v is either scalar or irreducible, π_v must be a twist of an unramified principal series. By local–global compatibility, this implies that a twist of $\rho|_{G_{F_v}}$ is unramified and pure (that is, satisfies the Ramanujan conjecture) by [**Bla06**, Theorem 1]. But this contradicts the assumption that $\rho|_{G_{F_v}}$ arises from a point of Spec $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}}$ (see the description in 5.2.6).

Finally, local–global compatibility for D implies that π^D corresponds to an eigenform in $S_\tau(U, \mathcal{O})$. □

5.4. Patching

5.4.1. Let $R^{\square,\psi}_{F,S}$ denote the complete local $\mathcal{O}$-algebra which prorepresents the functor which assigns to a local Artinian $\mathcal{O}$-algebra A the set of equivalence classes of tuples $(\rho, \{\alpha_v\}_{v\in S})$, where ρ is a lifting of $\overline{\rho}$ as a $G_{F,S}$-representation with determinant $\psi\varepsilon^{-1}$, $\alpha_v \in \ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(\mathbb{F}))$, and two such tuples $(\rho, \{\alpha_v\}_{v\in S})$ and $(\rho', \{\alpha'_v\}_{v\in S})$ are equivalent if there is an element $\beta \in \ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(\mathbb{F}))$ with $\rho' = \beta\rho\beta^{-1}$ and $\alpha'_v = \beta\alpha_v$ for all $v \in S$, so that $R^{\square,\psi}_{F,S}$ is naturally an R^{ψ}_S-algebra. We define $R^{\square,\tau,\psi}_{F,S} = R^{\square,\psi}_{F,S} \otimes_{R^{\psi}_S} R^{\tau,\psi}_S$. We also have the corresponding universal deformation ring $R^{\tau,\psi}_{F,S}$, defined as the image of $R^{\psi}_{F,S}$ in $R^{\square,\tau,\psi}_{F,S}$.

Choose a lift $\rho^{\mathrm{univ}}_S : G_{F,S} \to \mathrm{GL}_2(R^{\tau,\psi}_{F,S})$ representing the universal deformation. Fix a place $w \in S$, and let

$$\mathcal{T} = \mathcal{O}[[X_{v,i,j} : v \in S, i, j = 1, 2]]/(X_{w,1,1}).$$

The tuple $(\rho^{\mathrm{univ}}_S, (1_2 + X_{v,i,j})_{v\in S})$ induces an isomorphism $R^{\square,\tau,\psi}_{F,S} \xrightarrow{\sim} R^{\tau,\psi}_{F,S}\widehat{\otimes}_{\mathcal{O}}\mathcal{T}$. (Note that $1_2 + X_{v,i,j}$ is a 2×2 matrix, and the fact that $X_{w,1,1} = 0$ in $\mathcal{T}$ implies that this tuple has no nontrivial scalar endomorphisms.)

We let $M = S_\tau(U, \mathcal{O})_{\mathfrak{m}}$. Fix a filtration by $\mathbb{F}$-subspaces

$$0 = L_0 \subset L_1 \subset \cdots \subset L_s = (\otimes_{v|p,\mathcal{O}} L_{\tau_v}) \otimes_{\mathcal{O}} \mathbb{F}$$

such that each L_i is $\prod_{v|p} U_v$-stable, and, for each $i = 0, 1, \ldots, s-1$, $\sigma_i := L_{i+1}/L_i$ is absolutely irreducible. This in turn induces a filtration

$$0 = M^0 \subset M^1 \subset \cdots \subset M^s = M \otimes_{\mathcal{O}} \mathbb{F}$$

on $M \otimes_{\mathcal{O}} \mathbb{F}$, whose graded pieces are the finite-dimensional $\mathbb{F}$-vector spaces $S_{\tau^p}(U, \sigma_i)_{\mathfrak{m}}$.

Let $q \geqslant [F : \mathbb{Q}]$ be an integer, set $g = q - [F : \mathbb{Q}] + |S| - 1$,

$$\begin{aligned} \Delta_\infty &= \mathbb{Z}_p^q, \\ R_\infty &= R_S^{\tau,\psi}[[x_1, \ldots, x_g]], \\ R'_\infty &= R_S^{\psi}[[x_1, \ldots, x_g]], \\ S_\infty &= \mathcal{T}[[\Delta_\infty]], \end{aligned}$$

and let $\mathfrak{a}$ denote the kernel of the $\mathcal{O}$-algebra homomorphism $S_\infty \to \mathcal{O}$ which sends each $X_{v,i,j}$ to zero and each element of Δ_∞ to one. Note that S_∞ is formally smooth over $\mathcal{O}$ of relative dimension $q + 4|S| - 1$ and that R_∞ also has relative dimension $q + 4|S| - 1$ over $\mathcal{O}$.

Now, a patching argument as in Section 4.3 shows for some $q \geqslant [F : \mathbb{Q}]$, there exist the following:

- an $\mathcal{O}$-module homomorphism $S_\infty \to R_\infty$, and an R_∞-module M_∞ which is finite free as an S_∞-module;
- a filtration by R_∞-modules

$$0 = M_\infty^0 \subset M_\infty^1 \subset \cdots \subset M_\infty^s = M_\infty \otimes \mathbb{F}$$

  whose graded pieces are finite free $S_\infty/\pi S_\infty$-modules;
- a surjection of $R_S^{\tau,\psi}$-algebras $R_\infty/\mathfrak{a}R_\infty \to R_{F,S}^{\tau,\psi}$; and
- an isomorphism of R_∞-modules $M_\infty/\mathfrak{a}M_\infty \xrightarrow{\sim} M$ which identifies M^i with $M_\infty^i/\mathfrak{a}M_\infty^i$.

As in Section 4.3, we may and do assume that, for $i = 1, 2, \ldots, s$, the (R'_∞, S_∞)-bimodule $M_\infty^i/M_\infty^{i-1}$ and the isomorphism $M_\infty^i/(\mathfrak{a}M_\infty^i + M_\infty^{i-1}) \xrightarrow{\sim} M^i/M^{i-1}$ depends only on $(U, \mathfrak{m}$, the types τ_v for $v \nmid p$ and) the isomorphism class of L_i/L_{i-1} as a $\prod_{v|p} U_v$-representation, but not on the choice of types τ_v for $v \mid p$. We denote this (R'_∞, S_∞)-bimodule by M_∞^σ, where $\sigma \xrightarrow{\sim} L_i/L_{i-1}$.

5.4.2. Assume that for $v \in S$, $v \nmid p$, the τ_v and the set $\widetilde{\Sigma' \text{ have}}$ been chosen such that $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}} \neq 0$, and that for $v \mid p$ we have $\det \tau_v = \varepsilon \det \overline{\rho}|_{G_{F_v}}$. We set

$$\mu'_\sigma(\overline{\rho}) = \frac{1}{2^m \prod_{v\in S, v\nmid p} e(R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}}/\pi)} e_{R_\infty/\pi}(M^\sigma_\infty).$$

We need the following variant of Lemma 4.3.9.

LEMMA 5.4.3. *We have the following.*

(1) *The support of M meets every irreducible component of* Spec $R^{\tau,\psi}_S[1/p]$.

(2) *$M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a finite $R_\infty[1/p]$-module, which which is locally free of rank 2^m over the formally smooth locus in* Spec $R_\infty[1/p]$.

(3) *$R^{\tau,\psi}_{F,S}$ is a finite $\mathcal{O}$-algebra and $M \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a faithful $R^{\tau,\psi}_{F,S}$-module.*

(4)
$$e(R_\infty/\pi R_\infty) = \sum_{i=1}^{s} 2^{-m} e_{R_\infty/\pi}(M^{\sigma_i}_\infty),$$

where σ_i is a global Serre weight with $L_i/L_{i-1} \xrightarrow{\sim} \sigma_i$.

Proof. (1) follows from Lemma 5.3.6.

For (2), note that M_∞ is a finite R_∞-module, since it is a finite S_∞-module, and the same argument as in Lemma 4.3.9 shows that M_∞ is locally free over the formally smooth locus of Spec $R_\infty[1/p]$. Thus, to show (2), we have to show that any irreducible component of $Z \subset$ Spec $R_\infty[1/p]$ contains a closed point in the smooth locus of Spec $R_\infty[1/p]$ at which M_∞ has rank 2^m.

By (1), there exists $\mathfrak{p} \in Z$ in the support of M. We claim that the image of $\mathfrak{p}$ is in the smooth locus of Spec $R^{\tau,\psi}_S[1/p]$. Assuming this, we see that $\mathfrak{p}$ is in the smooth locus of Spec $R_\infty[1/p]$, and in particular M_∞ is locally free (of positive rank) at $\mathfrak{p}$. Hence the maps

$$R_\infty/\mathfrak{a}R_\infty[1/p] \to R^{\tau,\psi}_{F,S}[1/p] \to \mathbb{T}^S_\tau(U,\mathcal{O})_\mathfrak{m}[1/p]$$

become isomorphisms after localizing at $\mathfrak{p}$. Since $M \otimes \mathbb{Q}_p$ has rank 2^m over $\mathbb{T}^S_\tau(U,\mathcal{O})_\mathfrak{m}[1/p]$, it follows that M_∞ has rank 2^m at $\mathfrak{p}$.

By the description of $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}}$ given in 5.2.6, to show that the image of $\mathfrak{p}$ is in the smooth locus of Spec $R^{\tau,\psi}_S[1/p]$, it suffices to show that, if $v \in S\backslash\Sigma$, $v \nmid p$, and τ_v is scalar, then the image of $\mathfrak{p}$ lies on exactly one irreducible component of Spec $R^{\square,\tau_v,\psi}_{\overline{\rho}|_{G_{F_v}}}[1/p]$. Let $\rho_{\mathfrak{p}}$ denote the representation of $G_{F,S}$ corresponding to the image of $\mathfrak{p}$ in Spec $R^{\tau,\psi}_{F,S}$. Since $\mathfrak{p}$ is in the support of M, it corresponds to an automorphic representation π of $D^{\times}$, and π_v is a twist either of an unramified principal series, or of a Steinberg representation. In the former case, one sees as in the proof of Lemma 5.3.6 that $\rho_{\mathfrak{p}}|_{G_{F_v}}$ is the twist of an unramified representation which satisfies the Ramanujan conjecture. So $\mathfrak{p}$ is contained in a single component of Spec $R^{\square,\tau_v,\psi}_{\overline{\rho}|_{G_{F_v}}}[1/p]$, and this component parameterizes potentially unramified representations. When π_v is a twist of a Steinberg representation, then it follows from [**Car86**, Theorem A] that the image of $\rho_{\mathfrak{p}}|_{I_{F_v}}$ is infinite, and again $\mathfrak{p}$ lies on exactly one component of Spec $R^{\square,\tau_v,\psi}_{\overline{\rho}|_{G_{F_v}}}[1/p]$.

This proves (2), and also shows that $M\otimes\mathbb{Q}_p$ is a finite free $R_\infty/\mathfrak{a}R_\infty[1/p]$-module of rank 2^m. In particular, the map $R_\infty/\mathfrak{a}R_\infty[1/p]\to R^{\tau,\psi}_{F,S}[1/p]$ is an isomorphism, so $M\otimes\mathbb{Q}_p$ is a finite free $R^{\tau,\psi}_{F,S}[1/p]$-module, which proves (3).

Finally, by (2), M_∞ is locally free of rank 2^m at the generic points of R_∞. By [**Kis09a**, Lemma 1.3.4], this implies that $e(R_\infty/\pi)=2^{-m}e_{R_\infty/\pi}(M_\infty/\pi M_\infty)$, and (4) follows as in the proof of Lemma 4.3.9. □

PROPOSITION 5.4.4. *For each global Serre weight σ, we have $\mu'_\sigma(\overline{\rho})=\prod_{v|p}\mu_{\sigma_v}(\overline{\rho}|_{G_{F_v}})$, where $\sigma\cong\otimes_{v|p}\sigma_v$, and the $\mu_{\sigma_v}(\overline{\rho}|_{G_{F_v}})$ are as in Theorem 4.5.5. In particular, $\mu'_\sigma(\overline{\rho})$ is a nonnegative integer.*

Proof. By Lemma 5.4.3, we have

$$e(R_\infty/\pi R_\infty)=2^{-m}\sum_{i=1}^{s}e_{R_\infty/\pi}(M_\infty^{\sigma_i}),$$

where σ_i is a global Serre weight with $L_i/L_{i-1}\xrightarrow{\sim}\sigma_i$. Since

$$e(R_\infty/\pi R_\infty)=e(R_S^{\tau,\psi}/\pi)=\prod_{v|p}e\big(R^{\square,0,\tau_v,\psi}_{\overline{\rho}|_{G_{F_v}}}/\pi\big)\prod_{v\in S,v\nmid p}e\big(R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}}/\pi\big),$$

this yields

$$\prod_{v|p}e\big(R^{\square,0,\tau_v,\psi}_{\overline{\rho}|_{G_{F_v}}}/\pi\big)=\sum_{i=1}^{s}\mu'_{\sigma_i}(\overline{\rho}),$$

which by Remark 5.2.2 is equivalent to

$$\prod_{v|p}e\big(R^{\square,0,\tau_v}_{\overline{\rho}|_{G_{F_v}}}/\pi\big)=\sum_{i=1}^{s}\mu'_{\sigma_i}(\overline{\rho}).$$

Since the σ_i are precisely the Jordan–Hölder factors of $(\otimes_{v|p,\mathcal{O}} L_{\tau_v}) \otimes_{\mathcal{O}} \mathbb{F}$, we see that this is equivalent to

$$\prod_{v|p} e\big(R^{\square,0,\tau_v}_{\overline{\rho}|G_{F_v}}/\pi\big) = \sum_{\{\sigma_v\}_{v|p}} \Big(\prod_{v|p} n_{0,\tau_v}(\sigma_v)\Big)\mu'_\sigma(\overline{\rho}),$$

where the sum runs over tuples $\{\sigma_v\}_{v|p}$ of equivalence classes of Serre weights, and $\sigma \cong \otimes_{v|p}\sigma_v$. By the above construction, these equations hold for all choices of types τ_v with $\det\tau_v = \varepsilon\,\widetilde{\det\overline{\rho}}|_{I_{F_v}}$.

In fact, we claim that these equations hold for all choices of types τ_v with tame determinant. In order to see this, suppose that, for some v, $\det\tau_v$ is tame, but $\det\tau_v \neq \varepsilon\,\widetilde{\det\overline{\rho}}|_{I_{F_v}}$, and so $\det\overline{\tau}_v \neq \overline{\varepsilon}\det\overline{\rho}|_{I_{F_v}}$. Then $R^{\square,0,\tau_v}_{\overline{\rho}|_{I_{F_v}}} = 0$, so it suffices to prove that $\mu'_\sigma(\overline{\rho}) = 0$ whenever $n_{0,\tau_v}(\sigma_v) \neq 0$ for all $v \mid p$ (as then the equation will collapse to $0 = 0$). Equivalently, we must show that $M^\sigma_\infty = 0$ if $n_{0,\tau_v}(\sigma_v) \neq 0$; but, as in Corollary 4.5.3, this is an easy consequence of local–global compatibility and the assumption that $\det\overline{\tau}_v \neq \overline{\varepsilon}\det\overline{\rho}|_{I_{F_v}}$, since if $M^\sigma_\infty \neq 0$ then $M^\sigma \neq 0$.

Now, by Lemma 4.5.1(1) and Lemma 4.5.2, we see that the quantities $\mu'_\sigma(\overline{\rho})$ are uniquely determined by these equations, as the τ_v run over all types with tame determinant. However, by Corollary 4.5.6, we see that setting $\mu'_\sigma(\overline{\rho}) = \prod_{v|p}\mu_{\sigma_v}(\overline{\rho}|G_{F_v})$ gives a solution to the equations, so we must have $\mu'_\sigma(\overline{\rho}) = \prod_{v|p}\mu_{\sigma_v}(\overline{\rho}|G_{F_v})$, as required. □

The following corollary is the main result we need to apply our techniques to the weight part of Serre's conjecture.

COROLLARY 5.4.5. *With the above notation and assumptions, let $\sigma = \otimes_{v|p}\sigma_v$ be a global Serre weight. Then we have $S_{\tau^p}(U,\sigma)_{\mathfrak{m}} \neq 0$ if and only if $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|G_{F_v})$ for all places $v \mid p$.*

Proof. By Proposition 5.4.4, we see that $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|G_{F_v})$ for all places $v \mid p$ if and only if $e_{R_\infty/\pi}(M^\sigma_\infty) \neq 0$. Since (by the patching construction) M^σ_∞ is maximal Cohen–Macaulay over R_∞/π, we see from [Mat89, Theorem 13.4] that $e_{R_\infty/\pi}(M^\sigma_\infty) \neq 0$ if and only if $M^\sigma_\infty \neq 0$; but $M^\sigma_\infty \neq 0$ if and only if $M^\sigma \neq 0$, and $M^\sigma = S_{\tau^p}(U,\sigma)_{\mathfrak{m}}$ by definition. □

5.5. Proof of the BDJ conjecture. We are now ready to prove Theorem B, which amounts to translating Corollary 5.4.5 into the original formulation of the BDJ conjecture.

5.5.1. Let D and ψ be as above and, as usual, denote by Σ the set of finite places where D is ramified. Let $V = (\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v}))V^p$ with V^p a compact open subgroup of $(D \otimes_{\mathbb{Q}} \mathbb{A}^{\infty,p})^\times$. We assume that $\psi \circ \det|_{V^p}$ is trivial. Let $\sigma = \otimes_{v|p}\sigma_v$ be a global Serre weight for D. We assume that σ is compatible with ψ as above, and extend σ to a representation of $V(\mathbb{A}_F^\infty)^\times$. Let τ_0^p denote the trivial $\mathcal{O}$-representation of V^p. We write $S(V, \sigma) = S_{\tau_0^p}(V, \sigma)$, and we remind the reader that the definition of $S(V, \sigma)$ depends on ψ.

Let T be a finite set of finite places of F, containing the set of places where D, $\overline{\rho}$, or ψ are ramified and all the places dividing p, and such that $(\mathcal{O}_D)_v^\times \subset V$ for $v \notin T$. We again denote by $\mathfrak{m} \subset \mathbb{T}^{T,\mathrm{univ}}$ the maximal ideal associated to $\overline{\rho}$, as above. The Hecke algebra $\mathbb{T}^{T,\mathrm{univ}}$ acts on $S(V, \sigma)$, and we may consider the localization $S(V, \sigma)_\mathfrak{m}$. Note that $S(V, \sigma)_\mathfrak{m}$ does not depend on the choice of the set T satisfying the above conditions. Thus, we may write $S(V, \sigma)_\mathfrak{m}$ without making a choice of T.

DEFINITION 5.5.2. We say that $\overline{\rho}$ is modular for D of weight σ if, for some V as above, we have $S(V, \sigma)_\mathfrak{m} \neq 0$.

DEFINITION 5.5.3. We say that $\overline{\rho}$ is *compatible with* D if, for all $v \in \Sigma$, $\overline{\rho}|_{G_{F_v}}$ is either irreducible, or is a twist of an extension of the trivial character by the cyclotomic character.

COROLLARY 5.5.4. *Let $p > 2$ be prime, let F be a totally real field, and let $\overline{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Assume that $\overline{\rho}$ is modular, that $\overline{\rho}|_{G_{F(\zeta_p)}}$ is irreducible, and if $p = 5$ assume further that the projective image of $\overline{\rho}|_{G_{F(\zeta_p)}}$ is not isomorphic to A_5.*

For each place $v \mid p$ of F with residue field k_v, let σ_v be a Serre weight of $\mathrm{GL}_2(k_v)$. Then $\overline{\rho}$ is modular for D of weight $\sigma = \otimes_{v|p}\sigma_v$ if and only if $\overline{\rho}$ is compatible with D and $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all $v \mid p$.

Proof. Suppose first that $\overline{\rho}$ is compatible with D and that $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all $v \mid p$. Let S be a set of primes containing all the primes dividing $v \mid p$, and all primes where $\overline{\rho}$, $\bar{\psi}$, or D ramifies. We take the set $\Sigma' \subset S$ to be empty. Since $\overline{\rho}$ is compatible with D, for each $v \in \Sigma$, $\overline{\rho}|_{G_{F_v}}$ has a lift of discrete series type, as in [BD13, Proof of Corollaire 3.2.3], and we can choose our types τ_v for $v \in S$, $v \nmid p$, such that $R^{\square,\tau_v,\psi,*}_{\overline{\rho}|_{G_{F_v}}} \neq 0$.

Choose a compact open subgroup $U \subset (D \otimes_{\mathbb{Q}} \mathbb{A}^\infty)^\times$, and the lattices $L_{\tau_v} \subset \sigma(\tau_v)$ as in 5.3 above. Let $V = (\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v}))V^p \subset U$ be a normal subgroup such that V acts trivially on $\sigma(\tau_v)$ for all $v \in S$, $v \nmid p$, and $\psi \circ \det|_{V^p}$ is trivial.

Then

$$S_{\tau^p}(U,\sigma)_{\mathfrak{m}} = S_{\tau^p}(V,\sigma)_{\mathfrak{m}}^{U/V} = (S(V,\sigma)_{\mathfrak{m}} \otimes_{\mathcal{O}} L_{\tau^p})^{U/V}. \tag{5.5.5}$$

Since $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all $v \mid p$, we have $S_{\tau^p}(U,\sigma)_{\mathfrak{m}} \neq 0$ by Corollary 5.4.5, and $S(V,\sigma)_{\mathfrak{m}} \neq 0$, as required.

Conversely, suppose that $S(V,\sigma)_{\mathfrak{m}} \neq 0$ for some V. Let S be a set of primes containing the primes $v \mid p$, the primes where $\overline{\rho}$, $\bar{\psi}$, or D ramify, and the primes v such that $(\mathcal{O}_D)_v^\times$ is not contained in V. Let $S^p = \{v \in S : v \nmid p\}$. Write $V^{S^p} = \prod_{v\notin S^p} V_v$, and set

$$Y = \varinjlim_{W_{S^p}} S(W_{S^p}V^{S^p},\sigma)_{\mathfrak{m}} \otimes_{\mathbb{F}} \bar{\mathbb{F}}_p,$$

where W_{S^p} runs over compact open subgroups of $\prod_{v\in S^p} V_v$. Then Y is naturally a smooth $\bar{\mathbb{F}}_p$-representation of the group $D_{S^p}^\times = \prod_{v\in S^p} D_v^\times$. Note that $Y \neq 0$ as $S(V,\sigma)_{\mathfrak{m}} \neq 0$. Let $\bar{\pi} \subset Y$ be an irreducible subrepresentation. Then $\bar{\pi} = \otimes_{v\in S^p}\bar{\pi}_v$, where each $\bar{\pi}_v$ is an irreducible, admissible, smooth representation of $D_v^\times$. Let $\Sigma' \subset S^p$ be the subset where D is unramified at v and $\bar{\pi}_v$ is finite dimensional.

We now define the $(\mathcal{O}_D)_v^\times$-representations L_{τ_v}, for $v \in S^p$, used in the definition of our spaces of modular forms. If either $\bar{\pi}_v$ is infinite dimensional or D is ramified at v, then we take τ_v and L_{τ_v} to be the type and $(\mathcal{O}_D)_v^\times$-representation produced by applying Lemmas 5.1.1 and 5.1.3 respectively to the representation $\bar{\pi}_v$. If D is unramified at v, and $\bar{\pi}_v$ is finite dimensional, then, by [**Vig89b**, Theorem 1.1], $\bar{\pi}_v$ has the form $\bar{\chi}_v \circ \det$, where $\bar{\chi}_v : F_v^\times \to \bar{\mathbb{F}}_p^\times$ is a continuous character. Since $\bar{\chi}_v$ has finite image, we may lift $\bar{\chi}_v|_{\mathcal{O}_{F_v}^\times}$ to a character $\chi_v : \mathcal{O}_{F_v}^\times \to \bar{\mathbb{Z}}_p^\times$. We take τ_v to be the two-dimensional scalar representation given by $\chi_v \circ \mathrm{Art}_{F_v}^{-1}$, and $L_{\tau_v} = \mathcal{O}$ with $(\mathcal{O}_D)_v^\times$ acting via $\chi_v^{-1} \circ \det$. Since Y has central character $\bar{\psi}$, so does $\bar{\pi}$ and each $\bar{\pi}_v$ for $v \in S^p$. Thus $\det \tau_v$ reduces to $\bar{\psi}$ mod p. Since $p \neq 2$, after replacing each τ_v by a twist by a character which has trivial reduction, we may assume that $\det \tau_v = \psi|_{I_{F_v}}$ for $v \in S^p$. We take $\Sigma' \subset S^p$ to be the set of primes where D is unramified and $\bar{\pi}_v$ is finite dimensional.

Set $\mathcal{O}_{D_{S^p}^\times} = \prod_{v\in S^p}(\mathcal{O}_D)_v^\times$. By construction, $(Y \otimes L_{\tau^p})^{\mathcal{O}_{D_{S^p}^\times}} \neq 0$. Hence there is an open normal subgroup $W_{S^p} \subset \mathcal{O}_{D_{S^p}^\times}$, such that $(S(W_{S^p}V^{S^p},\sigma)_{\mathfrak{m}} \otimes L_{\tau^p})^{\mathcal{O}_{D_{S^p}^\times}} \neq 0$. Let $U = \prod_v U_v$ be as in 5.3, so in particular $U_v \subset (\mathcal{O}_D)_v^\times$. By what we just saw, there exists a normal open subgroup $W = (\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v}))W^p \subset U$ such that W acts trivially on $\sigma(\tau_v)$ for $v \in S^p\backslash\Sigma'$, $\tau_v \circ \mathrm{Art}_{F_v}$ is trivial on W if $v \in \Sigma'$, and $\psi \circ \det|_{W^p}$ is trivial, and such that $(S(W,\sigma)_{\mathfrak{m}} \otimes L_{\tau^p})^{U/W} \neq 0$. Then $S_{\tau^p}(U,\sigma)_{\mathfrak{m}} \neq 0$ by (5.5.5), applied with W in place of V. This implies that $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all $v \mid p$, by Corollary 5.4.5.

It remains to see that $\bar{\rho}$ is compatible with D. Choose a representation of $\prod_{v|p}(\mathcal{O}_D)_v^\times$ on a finite free $\mathcal{O}$-module θ such that σ is a Jordan–Hölder factor of $\theta \otimes_{\mathcal{O}} \mathbb{F}$. Then $S_{\tau^p}(U, \sigma)_\mathfrak{m} \neq 0$ implies that $S_{\tau^p}(U, \theta)_\mathfrak{m} \neq 0$. This implies that, for $v \in \Sigma$, $\bar{\rho}|_{G_{F_v}}$ has a lift of discrete series type, and hence that $\bar{\rho}$ is compatible with D, as in [BD13, Proof of Corollaire 3.2.3]. □

REMARK 5.5.6. The above arguments strongly suggest that there should be versions of the Breuil–Mézard conjecture for quaternion algebras, and also for mod l representations of the absolute Galois groups of p-adic fields, with $l \neq p$. The first problem is considered in [GG13], and the second problem in [Sho13].

Following [BDJ10], we say that $\bar{\rho}$ is modular of weight σ if there exists a quaternion algebra D over F which is indefinite at a single prime $v_0|\infty$ and split at all primes dividing p, and such that $\overline{\rho}$ is modular for D of weight σ. (In the definition of [BDJ10], the prime v_0 is fixed. However, it follows from the Jacquet–Langlands correspondence that the condition does not in fact depend on the choice of v_0.) As a consequence of Corollary 5.5.4 we have the following result on the BDJ conjecture as formulated in [BDJ10, Conjecture 3.14].

COROLLARY 5.5.7 (Theorem B). *Let p, F, $\overline{\rho}$, and σ be as in 5.5.4. Then $\overline{\rho}$ is modular of weight σ if and only if $\sigma_v \in W^{\mathrm{BT}}(\overline{\rho}|_{G_{F_v}})$ for all $v \mid p$.*

Proof. The 'only if' direction follows immediately from Corollary 5.5.4, which also implies the converse once we show that there is a D indefinite at a single prime $v_0|\infty$ and split at all primes dividing p such that $\overline{\rho}$ is compatible with D. For this, we may take D to be a quaternion algebra ramified at one infinite prime, and at one finite prime v, such that $\overline{\rho}(\mathrm{Frob}_v) = 1$ and $\mathbf{N}(v) = 1$ modulo p. □

Acknowledgements

The first author was partially supported by a Marie Curie Career Integration Grant, and by an ERC Starting Grant. The second author was partially supported by NSF grant DMS-0701123.

We would like to thank Frank Calegari, Fred Diamond, Matthew Emerton, Guy Henniart, Florian Herzig, and David Savitt for helpful conversations. We would also like to thank Matthew Emerton and Florian Herzig for their helpful comments on an earlier draft of this paper. T. G. would like to thank the mathematics department of Northwestern University for its hospitality in the final stages of this project.

Appendix A. Realizing local representations globally

In this appendix, we realize local representations globally, using the potential automorphy techniques of [BLGGT14a] and [Cal12]. We will freely use the notation and terminology of [BLGGT14a], in particular the notions of RACSDC, RAESDC, and RAECSDC automorphic representations π of $\mathrm{GL}_n(\mathbb{A}_{F^+})$ and $\mathrm{GL}_n(\mathbb{A}_F)$, and the associated p-adic Galois representations $r_{p,\iota}(\pi)$, which are defined in [BLGGT14a, Section 2.1].

PROPOSITION A.1. *Let $K/\mathbb{Q}_p$ be a finite extension, and let $\bar{r}_K : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is a totally real field L^+ and a continuous irreducible representation $\bar{r} : G_{L^+} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that the following hold.*

- *For each place $v \mid p$ of L^+, $L^+_v \cong K$ and $\bar{r}|_{L^+_v} \cong \bar{r}_K$.*
- *For each finite place $v \nmid p$ of L^+, $\bar{r}|_{G_{L^+_v}}$ is unramified.*
- *For each place $v \mid \infty$ of L^+, $\det \bar{r}(c_v) = -1$, where c_v is a complex conjugation at v.*
- *There is a nontrivial finite extension $\mathbb{F}/\mathbb{F}_p$ such that $\bar{r}(G_{L^+}) = \mathrm{GL}_2(\mathbb{F})$.*

Proof. Choose a nontrivial finite extension $\mathbb{F}/\mathbb{F}_p$ such that $\bar{r}_K(G_K) \subset \mathrm{GL}_2(\mathbb{F})$. We apply [Cal12, Proposition 3.2] where, in the notation of that result, the following hold.

- $G = \mathrm{GL}_2(\mathbb{F})$.
- We let $F = E$ be any totally real field with the property that, if $v \mid p$ is a place of E, then $E_v \cong K$.
- If $v \mid p$, we let $D_v = \bar{r}_K(G_K)$.
- If $v \mid \infty$, we let c_v be the nontrivial conjugacy class of order two in $\mathrm{GL}_2(\mathbb{F})$.

This produces a representation $\bar{r}$ which satisfies all the required properties, except possibly for the requirement that $\bar{r}$ be unramified outside p, which may be arranged by a further (solvable, if one wishes) base change. (Note that, while this is not stated there, the commutative diagram in of [Cal12, Proposition 3.2(4)] is the obvious one, where the top arrow is the natural inclusion.) □

We remark that the reason that we have assumed that $\mathbb{F} \neq \mathbb{F}_p$ is that we wish to apply [BLGG13b, Proposition A.2.1] to conclude that $\bar{r}(G_{L^+})$ is adequate. We have the following now standard potential modularity result.

THEOREM A.2. *Let L^+ be a totally real field, let M/L^+ be a finite extension, and let $p > 2$ be a prime. Let $\overline{\rho} : G_{L^+} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be an irreducible continuous totally odd representation. Then there exists a finite totally real Galois extension F^+/L^+ such that the following hold.*

- *The primes of L^+ above p split completely in F^+.*
- *F^+ is linearly disjoint from M over L^+.*
- *There is a weight zero RAESDC (regular, algebraic, essentially self-dual, cuspidal) automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_{F^+})$ such that $\bar{r}_{p,\iota}(\pi) \cong \overline{\rho}|_{G_{F^+}}$, and π has level potentially prime to p.*

Proof. This is a special case of [Sno09, Proposition 8.2.1], once one knows that, for each place $v \mid p$, $\overline{\rho}|_{G_{L_v^+}}$ admits a potentially Barsotti–Tate lift. (Such a lift will be of type A or B in the terminology of [Sno09].) If $\overline{\rho}|_{G_{L_v^+}}$ is irreducible, this is proved in the course of [Sno09, Proof of Proposition 7.8.1], and if it is reducible then it is an immediate consequence of [BLGG12, Lemma 6.1.6]. □

Combining Proposition A.1 and Theorem A.2, we obtain the following result.

COROLLARY A.3. *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be a finite extension, and let $\bar{r}_K : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is a totally real field F^+ and an RAESDC automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_{F^+})$ such that $\bar{r}_{p,\iota}(\pi)$ is absolutely irreducible, and such that the following hold.*

- *For each place $v \mid p$ of F^+, $F_v^+ \cong K$ and $\bar{r}_{p,\iota}(\pi)|_{F_v^+} \cong \bar{r}_K$.*
- *For each finite place $v \nmid p$ of F^+, $\bar{r}_{p,\iota}(\pi)|_{G_{F_v^+}}$ is unramified.*
- *π has level potentially prime to p.*
- *There is a nontrivial finite extension $\mathbb{F}/\mathbb{F}_p$ such that $\bar{r}_{p,\iota}(\pi)(G_{F^+}) = \mathrm{GL}_2(\mathbb{F})$.*

Proof. This follows at once by applying Theorem A.2 to the representation $\bar{r}$ provided by Proposition A.1, taking the auxiliary field M to be $(\overline{L})^{\ker \bar{r}}$. □

THEOREM A.4. *Suppose that $p > 2$, that $K/\mathbb{Q}_p$ is a finite extension, and let $\bar{r}_K : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is an imaginary CM field F with maximal totally real subfield F^+ and an RACSDC automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_F)$ such that $\bar{r}_{p,\iota}(\pi)$ is absolutely irreducible, and such that the following hold.*

- *Each place $v \mid p$ of F^+ splits in F.*
- *For each place $v \mid p$ of F^+, $F_v^+ \cong K$.*
- *For each place $v \mid p$ of F^+, there is a place $\widetilde{v}$ of F lying over v such that $\bar{r}_{p,\iota}(\pi)|_{G_{F_{\widetilde{v}}}}$ is isomorphic to an unramified twist of $\bar{r}_K$.*
- *$\bar{r}_{p,\iota}(\pi)$ is unramified outside of the places dividing p.*
- *$\zeta_p \notin F$.*
- *$\bar{r}_{p,\iota}(\pi)(G_{F(\zeta_p)})$ is adequate.*
- *The projective image of $\bar{r}$ is not isomorphic to A_4.*

Proof. Let π be the RAESDC automorphic representation of $\mathrm{GL}_2(\mathbb{A}_{F^+})$ provided by Corollary A.3. Choose a totally imaginary quadratic extension F/F^+ which is linearly disjoint from $(\overline{F})^{\ker \bar{r}_{p,\iota}(\pi)}(\zeta_p)$ over F and in which all places of F^+ lying over p split. For each place $v \mid p$ of F^+, choose a place $\widetilde{v}$ of F lying over v. Choose a finite order character $\theta : G_F \to \overline{\mathbb{Q}}_p^\times$ such that, for each place $v \mid p$ of F^+, $\theta|_{G_{F_{\widetilde{v}}}} = 1$, and $\theta|_{G_{F_{\widetilde{v}^c}}} = \varepsilon \det r_{p,\iota}(\pi)|_{G_{F_v^+}}$. (The existence of such a character is guaranteed by [CHT08, Lemma 4.1.1].) Let θ_{F^+} denote θ composed with the transfer map $G_{F^+}^{\mathrm{ab}} \to G_F^{\mathrm{ab}}$, so that $\theta_{F^+}|_{G_F} = \theta\theta^c$.

Choose a finite order character $\psi : G_F \to \overline{\mathbb{Q}}_p^\times$ such that

$$\psi\psi^c = \varepsilon^{-1}(\theta_{F^+}(\det r_{p,\iota}(\pi))^{-1})|_{G_F}$$

and such that each $\psi|_{G_{F_{\widetilde{v}}}}$ is unramified. (The existence of such a character follows by applying [CHT08, Lemma 4.1.5] to $\varepsilon^{-1}(\theta_{G_{F^+}} \det \bar{r}_{p,\iota}(\pi))^{-1}|_{G_F}$, choosing the integers m_τ of [CHT08, Lemma 4.1.5] to be zero. Note that, by virtue of the choice of θ above, the character $\varepsilon^{-1}(\theta_{F^+}(\det r_{p,\iota}(\pi))^{-1})|_{G_F}$ is crystalline at all places of F dividing p.)

Then the representation $r := r_{p,\iota}(\pi)|_{G_F} \otimes \psi\theta^{-1}$ satisfies $r^c \cong r^\vee \varepsilon^{-1}$, so by base change there is an RACSDC automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F)$ satisfying all the required properties, except possibly the property that $\bar{r}_{p,\iota}(\pi)$ is unramified outside of p, which may be arranged by a further solvable base change. (Note that the claims that $\bar{r}_{p,\iota}(\pi)(G_{F(\zeta_p)})$ is adequate and that the projective image of $\bar{r}$ is not isomorphic to A_4 are an immediate consequence of Proposition 4.3.2 and the fact that $\bar{r}_{p,\iota}(\pi)(G_F) = \mathrm{GL}_2(\mathbb{F})$ for some nontrivial extension $\mathbb{F}/\mathbb{F}_p$.) □

COROLLARY A.5. *Suppose that $p > 2$, that $K/\mathbb{Q}_p$ is a finite extension, and let $\bar{r}_K : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is an imaginary*

CM field F and a continuous irreducible representation $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that the following hold.

- *Each place $v \mid p$ of F^+ splits in F, and has $F^+_v \cong K$.*
- *For each place $v \mid p$ of F^+, there is a place $\widetilde{v}$ of F lying over F^+ with $\bar{r}|_{G_{F_{\widetilde{v}}}}$ isomorphic to an unramified twist of $\bar{r}_K$.*
- *$[F^+ : \mathbb{Q}]$ is even.*
- *F/F^+ is unramified at all finite places.*
- *$\zeta_p \notin F$.*
- *$\bar{r}$ is unramified outside of p.*
- *$\bar{r}$ is automorphic in the sense of Section 3.2.*
- *$\bar{r}(G_{F(\zeta_p)})$ is adequate.*
- *The projective image of $\bar{r}$ is not isomorphic to A_4.*

Proof. This follows immediately from Theorem A.4 and the theory of base change between GL_2 and unitary groups; see [**BLGG13b**, Section 2] (note that we can make a finite solvable base change to ensure that $[F^+ : \mathbb{Q}]$ is even and F/F^+ is unramified at all finite places without affecting the other conditions). □

Appendix B. Errata for [Kis09a]

In this appendix, we give some errata for [**Kis09a**]. We are very grateful to Kevin Buzzard, Wansu Kim, Vytautas Paskunas, and Fabian Sander for pointing these out. Unless otherwise mentioned, all references are to [**Kis09a**], and we freely use the notation of that paper.

B.1. In (1.1.6), the multiplicity $\mu_{n,m}(\overline{\rho})$ when $\overline{\rho} \sim \left(\begin{smallmatrix} \mu_\lambda & * \\ 0 & \mu_\lambda \end{smallmatrix}\right) \otimes \omega^m$ (that is, $n = p - 2$ and $\lambda = \lambda'$) is not correctly defined in some cases. Namely, this multiplicity is defined to be one if (the class of) $*$ is nontrivial. If $*$ is trivial, the multiplicity is not defined explicitly, but there is a speculation that it is two. The actual multiplicities have been computed by Fabian Sander [**San12**], and are one when $*$ is nontrivial and ramified, two when $*$ is nontrivial and unramified, and four when $*$ is trivial. For the purposes of correcting the argument in [**Kis09a**], $\mu_{n,m}(\overline{\rho})$ should be defined implicitly whenever $n = p - 2$ and $\lambda = \lambda'$, as was done in the case when $*$ is trivial.

The mistake in the argument occurs in the last paragraph of (1.7.5), where it is claimed that 'we have just seen that (1.7.6) is a closed immersion'. The argument only shows that (1.7.6) is a homeomorphism onto its image, which is an isomorphism over generic points, but it can have some nonreduced fibres.

To correct this, the last claim in the statement of (1.7.5), regarding the formal smoothness of R^{ord}, should be deleted, and in the statement of (1.7.14) 'unless $\overline{\rho} \sim \left(\begin{smallmatrix} \mu_\lambda & 0 \\ 0 & \mu_{\lambda'} \end{smallmatrix}\right) \otimes \omega^m$' should be replaced by 'unless $\overline{\rho} \sim \left(\begin{smallmatrix} \mu_\lambda & * \\ 0 & \mu_{\lambda'} \end{smallmatrix}\right) \otimes \omega^m$ and either the class of $*$ is trivial, or $\lambda = \lambda'$.' This change has no effect on any of the arguments which follow and, in particular, does not effect the statement of the main theorems.

B.2. In the definition of the functor in Lemma (1.7.4), add the condition that $G_{\mathbb{Q}_p}$ acts on $L_A \otimes_A \mathbb{F}$ via ω_1.

B.3. In (2.2), add the condition that $\mathbf{N}(v) \neq -1(p)$ for $v \in \Sigma$. Otherwise, the character γ_v in (2.2.10) may fail to be unique. In the applications, one can reduce to this case by base change, and even assume that $\mathbf{N}(v) = 1(p)$ for $v \in \Sigma$.

B.4. The proof of Proposition (2.2.15) implicitly uses that, for $v \in \Sigma$, the rings $\bar{R}_v^{\square,\psi}/\pi$ are irreducible, and generically reduced. The proof of this is practically identical to the proof of Lemma (1.7.5), using the formal smoothness proved in [Kis09b, Lemma 2.6.3], and the fact that the representation at a generic point of Spec $\bar{R}_v^{\square,\psi}/\pi$ cannot be scalar.

B.5. Lemma (2.2.1) is not correct; there are some $\overline{\rho}$ with small image which provide counterexamples. The mistake is in the first line of the second paragraph of the proof, where it is asserted that if g' satisfies (2.2.2) then so does gg'. Replacing g by gg' does not change the left-hand side of (2.2.2), but may change the right-hand side.

This lemma is used to show that the conditions imposed on the compact open subgroup $U = \prod_v U_v$ and the set of primes where U_v is not maximal compact can be satisfied in the application to the main theorem (2.2.18). More precisely, the conditions are (2.1.2): that

$$(U(\mathbb{A}_F^f)^\times \cap tD^\times t^{-1})/F^\times = 1$$

for all $t \in (D \otimes_F \mathbb{A}_F^f)^\times$, and (2.2)(4): that if U_v is not maximal compact then $1 - \mathbf{N}(v) \in \mathbb{F}^\times$, and that the ratio of the eigenvalues $\overline{\rho}(\mathrm{Frob}_v)$ is not in $\{1, \mathbf{N}(v), \mathbf{N}(v)^{-1}\}$. We explain two ways to correct this, one which works when $p > 3$, and one which works in general.

If $p > 3$, one can require in all of Section 2 that p splits completely in F (which is in any case assumed after (2.2.10) and in the main theorem), drop the condition (2.1.2), and require that U_v is maximal compact for all v (so that (2.2)(4) is vacuous). Indeed (2.1.2) is needed only to show in (2.1.4) that $S_{\sigma,\psi}(U_\Delta, A)_\mathfrak{m}$ is a free $A[\Delta]$-module. The proof given there works provided that the finite group $(U(\mathbb{A}_F^f)^\times \cap tD^\times t^{-1})/F^\times$ has prime to p order. But this condition can fail only if $F(\zeta_p)$ is a quadratic extension of F, which implies that $p = 3$, since p splits completely in F.

To correct the argument for any p, maintain condition (2.1.2), but require in (2.2)(4) only that $1 - \mathbf{N}(v) \in \mathbb{F}^\times$, and that the ratio of the eigenvalues $\overline{\rho}(\mathrm{Frob}_v)$ is not $\mathbf{N}(v)^{\pm 1}$, which is the usual condition, for which the existence of v is guaranteed by [DDT97, Lemma 4.11]. Let $R \subset S\backslash\Sigma_p$ be the set of primes such that the eigenvalues of $\overline{\rho}(\mathrm{Frob}_v)$ are equal. Then $S_{\sigma,\psi}(U, \mathcal{O}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a $\mathbb{T}_{\sigma,\psi}(U)_\mathfrak{m}[1/p]$-module of rank $2^{|R|}$. (Alternatively, we could, as in the present paper, omit the operators U_{π_w}, $w \in S\backslash\Sigma_p$, from the definition of $\mathbb{T}_{\sigma,\psi}(U)_\mathfrak{m}$ in (2.1.5), in which case the rank would be $2^{|R|}$ with $R = S\backslash\Sigma_p$.) We now explain the modifications necessary to the proof which involve, as in the present paper, keeping track of some factors of $2^{|R|}$.

Replace Lemma (2.2.11) with the following.

LEMMA B.5.1. *Let $\mathfrak{p} \in \mathrm{Spec}\, \bar{R}_\infty$ be a prime in the support of M_∞. Then*

$$\dim_{\kappa(\mathfrak{p})} M_\infty \otimes_{\bar{R}_\infty} \kappa(\mathfrak{p}) \geqslant 2^{|R|},$$

with equality if $\mathfrak{p}$ is a minimal prime of $\bar{R}_\infty$. Moreover, the following conditions are equivalent.

(1) *M_∞ is a faithful $\bar{R}_\infty$-module.*

(2) *M_∞ is a faithful $\bar{R}_\infty$-module of rank $2^{|R|}$ at all generic points of R_∞.*

(3) $e(\bar{R}_\infty/\pi\bar{R}_\infty) = 2^{-|R|} e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi\bar{R}_\infty)$.

(4) $e(\bar{R}_\infty/\pi\bar{R}_\infty) \leqslant 2^{-|R|} e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi\bar{R}_\infty)$.

If these conditions hold, and $\rho : G_{F,S} \to \mathrm{GL}_2(\mathcal{O})$ is a deformation of $\bar{\rho}$ such that, for $v \in \Sigma_p$, $\rho|_{I_v}$ is an extension of γ_v by $\gamma_v(1)$ if $v \nmid p$, and $\rho|_{G_{F_v}}$ is potentially semistable of type (k, τ_v, ψ) if $v \mid p$, then ρ is modular, and arises from an eigenform in $S_{\sigma,\psi}(U, \mathcal{O}) \otimes_\mathcal{O} E$.

Proof. Let $\mathcal{O}[[\Delta_\infty]] = \mathcal{O}[[y_1, \ldots, y_{h+j}]]$, as in the proof of Lemma (2.2.11). Since M_∞ is finite flat over $\mathcal{O}[[\Delta_\infty]]$, its support in $\mathrm{Spec}\, \bar{R}_\infty$ is a union of irreducible components, and any such irreducible component $Z \subset \operatorname{Spec} \bar{R}_\infty$ surjects onto $\operatorname{Spec} \mathcal{O}[\![\Delta_\infty]\!]$.

To prove the first statement, it suffices to prove it in the case of minimal of the support of M_∞, which follows once we show that M_∞ is free of rank $2^{|R|}$ over the smooth locus of any Z as above. The argument for this is analogous to that in the Lemma 5.4.3 of the present paper. It suffices to show that a prime $\mathfrak{p} \in Z$ which maps to $(y_1, \ldots, y_{h+j})$ in $\operatorname{Spec} \mathcal{O}[\![\Delta_\infty]\!]$ is a smooth point of Z, or equivalently that $\mathfrak{p}$ maps to a smooth point of $\operatorname{Spec} \bar{R}_v^{\square,\psi}[1/p]$ for each $v \mid p$.

Let $\mathrm{WD}(\rho_{\mathfrak{p}}|_{G_{F_v}})$ denote the Weil–Deligne representation attached to $\rho_{\mathfrak{p}}|_{G_{F_v}}$, and let π be the automorphic representation of $D^\times$ corresponding to $\mathfrak{p}$. If π_v is a twist of an unramified principal series then $\mathrm{WD}(\rho_{\mathfrak{p}}|_{G_{F_v}})$ is a twist of a pure unramified representation, by [**Bla06**] and [**KM74**]. (Patrick Allen has pointed out that the smoothness in this case may be deduced from the fact that π, and hence π_v, has a Whittaker model, without using the Ramanujan conjecture.) If π_v is a twist of a Steinberg representation, then $\mathrm{WD}(\rho_{\mathfrak{p}}|_{G_{F_v}})$ satisfies the monodromy-weight conjecture by [**Sai09**]. The description of the rings $\bar{R}_v^{\square,\psi}$ in the appendix of [**Kis09a**] shows that this implies that $\mathfrak{p}$ maps to a smooth point of $\operatorname{Spec} \bar{R}_v^{\square,\psi}[1/p]$.

Now (1)–(4) and the final claim follow, as in the proof of (2.2.11). □

Replace Lemma (2.2.15) with the following.

LEMMA B.5.2. *The $\bar{R}_\infty$-module $M_\infty^i/M_\infty^{i-1}$ is nonzero if and only if for each $v \mid p$ we have $\mu_{n_{i,v},m_{i,v}}(\bar{\rho}|_{G_{F_v}}) \neq 0$. Suppose that this condition holds. Then, for any prime $\mathfrak{p} \in \operatorname{Spec} \bar{R}_\infty/\pi\bar{R}_\infty$ in the support of $M_\infty^i/M_\infty^{i-1}$, we have*

$$\dim_{\kappa(\mathfrak{p})} M_\infty^i/M_\infty^{i-1} \otimes_{\bar{R}_\infty} \kappa(\mathfrak{p}) \geqslant 2^{|R|}.$$

Moreover, if for each $v \mid p$ we have $\bar{\rho}|_{G_{F_v}} \nsim \begin{pmatrix} \omega\chi & * \\ 0 & \chi \end{pmatrix}$ for any character $\chi : G_{F_v} \to \mathbb{F}^\times$, then

$$2^{-|R|} e(M_\infty^i/M_\infty^{i-1}, \bar{R}_\infty/\pi\bar{R}_\infty) \geqslant e_\Sigma \prod_{v|p} \mu_{n_{i,v},m_{i,v}}(\bar{\rho}|_{G_{F_v}}) := e_{\Sigma_p},$$

where

$$e_\Sigma := \prod_{v\in\Sigma} e(\bar{R}_v^{\square,\psi}/\pi\bar{R}_v^{\square,\psi}).$$

Proof. The first claim is proved in Lemma (2.2.15), which also proves that the support of $M_\infty^i/M_\infty^{i-1}$ is all of $\operatorname{Spec} \bar{R}_\infty^i$. The third claim follows from the second, just as in the proof of Lemma (2.2.15).

To prove the second claim, we remark that, by [Gee11a, Section 4.6], there is a smooth irreducible E-representation (as always extending E if necessary) $\tilde{\sigma}_i$ of $\prod_{v|p} \mathrm{GL}_2(\kappa(v))$ such that σ_i is a Jordan–Hölder factor of the reduction of $\tilde{\sigma}$, and $S_{\tilde{\sigma}_i,\psi}(U,\mathcal{O})_{\mathfrak{m}} \otimes \mathbb{F} \xrightarrow{\sim} S_{\sigma_i,\psi}(U,\mathcal{O})_{\mathfrak{m}}$ (see Lemma 4.6.3, and see the proof of 4.6.5, 4.6.6 of [Gee11a]). (The use of these auxiliary smooth representations can be avoided by working with unitary groups, as in the present paper.) Now, the second claim follows by applying Lemma B.5.1 with $\sigma = \tilde{\sigma}_i$. □

Finally, in the proofs of Corollary (2.2.17) and Lemma (2.3.1), all the expressions $e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi\bar{R}_\infty)$ and $e(M^i_\infty/\pi M^i_\infty, \bar{R}_\infty/\pi\bar{R}_\infty)$ should be multiplied by $2^{-|R|}$.

References

[BLGG11] T. Barnet-Lamb, T. Gee and D. Geraghty, 'The Sato–Tate conjecture for Hilbert modular forms', *J. Amer. Math. Soc.* **24**(2) (2011), 411–469.

[BLGG12] T. Barnet-Lamb, T. Gee and D. Geraghty, 'Congruences between Hilbert modular forms: constructing ordinary lifts', *Duke Math. J.* **161**(8) (2012), 1521–1580.

[BLGG13a] T. Barnet-Lamb, T. Gee and D. Geraghty, 'Congruences between Hilbert modular forms: constructing ordinary lifts, II', *Math. Res. Lett.* **20**(1) (2013), 67–72.

[BLGG13b] T. Barnet-Lamb, T. Gee and D. Geraghty, 'Serre weights for rank two unitary groups', *Math. Ann.* **356**(4) (2013), 1551–1598.

[BLGHT11] T. Barnet-Lamb, D. Geraghty, M. Harris and R. Taylor, 'A family of Calabi–Yau varieties and potential automorphy II', *Publ. Res. Inst. Math. Sci.* **47**(1) (2011), 29–98.

[BLGGT14a] T. Barnet-Lamb, T. Gee, D. Geraghty and R. Taylor, 'Potential automorphy and change of weight', *Ann. of Math. (2)* **179**(2) (2014), 501–609.

[BLGGT14b] T. Barnet-Lamb, T. Gee, D. Geraghty and R. Taylor, 'Local-global compatibility for $l = p$, II', *Ann. Sci. Éc. Norm. Supér.* **47**(1) (2014), 165–179.

[BC11] J. Bellaïche and G. Chenevier, 'The sign of Galois representations attached to automorphic forms for unitary groups', *Compositio Math.* **147**(5) (2011), 1337–1352.

[Bla06] D. Blasius, 'Hilbert modular forms and the Ramanujan conjecture', in *Noncommutative Geometry and Number Theory*, Aspects of Mathematics, E37 (Vieweg, Wiesbaden, 2006), 35–56.

[BD13] C. Breuil and F. Diamond, 'Formes modulaires de Hilbert modulo p et valeurs d'extensions Galoisiennes', *Ann. Sci. Éc. Norm. Supér.* (2014), to appear.

[BM02] C. Breuil and A. Mézard, 'Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbf{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ en $l = p$', *Duke Math. J.* **115**(2) (2002), 205–310; with an appendix by G. Henniart.

[BDJ10] K. Buzzard, F. Diamond and F. Jarvis, 'On Serre's conjecture for mod l Galois representations over totally real fields', *Duke Math. J.* **155**(1) (2010), 105–161.

[Cal12] F. Calegari, 'Even Galois representations and the Fontaine–Mazur conjecture II', *J. Amer. Math. Soc.* **25**(2) (2012), 533–554.

[Car86] H. Carayol, 'Sur les représentations l-adiques associées aux formes modulaires de Hilbert', *Ann. Sci. Éc. Norm. Supér. (4)* **19**(3) (1986), 409–468.

[CHT08] L. Clozel, M. Harris and R. Taylor, 'Automorphy for some l-adic lifts of automorphic mod l Galois representations', *Publ. Math. Inst. Hautes Études Sci.* **108** (2008), 1–181.

[DDT97] H. Darmon, F. Diamond and R. Taylor, 'Fermat's last theorem', in *Elliptic Curves, Modular Forms & Fermat's Last Theorem (Hong Kong, 1993)* (International Press, Cambridge, MA, 1997), 2–140.

[Dia07] F. Diamond, 'A correspondence between representations of local Galois groups and Lie-type groups', in *L-Functions and Galois Representations*, London Mathematical Society Lecture Note Series, 320 (Cambridge University Press, Cambridge, 2007), 187–206.

[EG14] M. Emerton and T. Gee, 'A geometric perspective on the Breuil–Mézard conjecture', *J. Inst. Math. Jussieu* **13**(1) (2014), 183–223.

[GL12] H. Gao and T. Liu, 'A note on potential diagonalizability of crystalline representations', *Math. Ann.* **360**(1–2) (2014), 481–487.

[Gee06] T. Gee, 'A modularity lifting theorem for weight two Hilbert modular forms', *Math. Res. Lett.* **13**(5–6) (2006), 805–811.

[Gee11a] T. Gee, 'Automorphic lifts of prescribed types', *Math. Ann.* **350**(1) (2011), 107–144.

[Gee11b] T. Gee, 'On the weights of mod p Hilbert modular forms', *Invent. Math.* **184**(1) (2011), 1–46.

[GG12] T. Gee and D. Geraghty, 'Companion forms for unitary and symplectic groups', *Duke Math. J.* **161**(2) (2012), 247–303.

[GG13] T. Gee and D. Geraghty, 'The Breuil–Mézard conjecture for quaternion algebras', (2013).

[GLS12] T. Gee, T. Liu and D. Savitt, 'Crystalline extensions and the weight part of Serre's conjecture', *Algebra Number Theory* **6**(7) (2012), 1537–1559.

[GLS13] T. Gee, T. Liu and D. Savitt, 'The weight part of Serre's conjecture for GL(2)', (2013).

[GLS14] T. Gee, T. Liu and D. Savitt, 'The Buzzard–Diamond–Jarvis conjecture for unitary groups', *J. Amer. Math. Soc.* **27**(2) (2014), 389–435.

[HT01] M. Harris and R. Taylor, 'The geometry and cohomology of some simple Shimura varieties', in *Annals of Mathematics Studies, Vol. 151* (Princeton University Press, Princeton, NJ, 2001); with an appendix by V. G. Berkovich.

[KM74] N. M. Katz and W. Messing, 'Some consequences of the Riemann hypothesis for varieties over finite fields', *Invent. Math.* **23** (1974), 73–77.

[KW09] C. Khare and J.-P. Wintenberger, 'On Serre's conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$', *Ann. of Math. (2)* **169**(1) (2009), 229–253.

[Kis08] M. Kisin, 'Potentially semi-stable deformation rings', *J. Amer. Math. Soc.* **21**(2) (2008), 513–546.

[Kis09a] M. Kisin, 'The Fontaine–Mazur conjecture for GL_2', *J. Amer. Math. Soc.* **22**(3) (2009), 641–690.

[Kis09b] M. Kisin, 'Moduli of finite flat group schemes, and modularity', *Ann. of Math. (2)* **170**(3) (2009), 1085–1180.

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

[Lab11] J.-P. Labesse, 'Changement de base C et séries discrètes', in *On the Stabilization of the Trace formula*, Stab. Trace Formula Shimura Var. Arith. Appl., 1 (Int. Press, Somerville, MA, 2011), 429–470.

[Mat89] H. Matsumura, 'Commutative ring theory', in *Cambridge Studies in Advanced Mathematics*, 2nd edn, Vol. 8 (Cambridge University Press, Cambridge, 1989), ; translated from the Japanese by M. Reid.

[New13] J. Newton, 'Serre weights and Shimura curves', *Proc. Lond. Math. Soc. (3)* **108**(6) (2014), 1471–1500.

[Pil08] V. Pilloni, 'The study of 2-dimensional p-adic Galois deformations in the $l \neq p$ case', (2008).

[Sai09] T. Saito, 'Hilbert modular forms and p-adic Hodge theory', *Compositio Math.* **145**(5) (2009), 1081–1113.

[San12] F. Sander, 'Hilbert–Samuel multiplicities of certain deformation rings', (2012).

[Sch08] M. M. Schein, 'Weights in Serre's conjecture for Hilbert modular forms: the ramified case', *Israel J. Math.* **166** (2008), 369–391.

[Ser77] J.-P. Serre, 'Linear representations of finite groups', in *Graduate Texts in Mathematics, Vol. 42* (Springer-Verlag, New York, 1977); translated from the second French edition by L. L. Scott.

[Sho13] J. Shotton, 'Local deformation rings and a Breuil–Mézard conjecture when $l \neq p$', (2013).

[Sno09] A. Snowden, 'On two dimensional weight two odd representations of totally real fields', (2009).

[Tay06] R. Taylor, 'On the meromorphic continuation of degree two L-functions', *Doc. Math.* (2006), 729–779; no. extra vol. (electronic).

[Tho12] J. Thorne, 'On the automorphy of l-adic Galois representations with small residual image', *J. Inst. Math. Jussieu* **11**(4) (2012), 855–920; with an appendix by R. Guralnick, F. Herzig, R. Taylor and Thorne.

[Tit66] J. Tits, 'Classification of algebraic semisimple groups', in *Algebraic Groups and Discontinuous Subgroups (Proceedings of Symposia in Pure Mathematics, Boulder, CO, 1965)* (American Mathematical Society, Providence, RI, 1966), 33–62.

[Vig89a] M.-F. Vignéras, 'Correspondance modulaire galois-quaternions pour un corps p-adique', in *Number Theory (Ulm, 1987)*, Lecture Notes in Mathematics, 1380 (Springer, New York, 1989), 254–266.

[Vig89b] M.-F. Vignéras, 'Représentations modulaires de GL$(2, F)$ en caractéristique l, F corps p-adique, $p \neq l$', *Compositio Math.* **72**(1) (1989), 33–66.