A GEOMETRIC PERSPECTIVE ON THE BREUIL–MÉZARD CONJECTURE

MATTHEW EMERTON AND TOBY GEE

Abstract. Let $p > 2$ be prime. We state and prove (under mild hypotheses on the residual representation) a geometric refinement of the Breuil–Mézard conjecture for 2-dimensional mod p representations of the absolute Galois group of $\mathbb{Q}_p$. We also state a conjectural generalisation to n-dimensional representations of the absolute Galois group of an arbitrary finite extension of $\mathbb{Q}_p$, and give a conditional proof of this conjecture, subject to a certain $R = \mathbb{T}$-type theorem together with a strong version of the weight part of Serre's conjecture for rank n unitary groups. We deduce an unconditional result in the case of two-dimensional potentially Barsotti–Tate representations.

1. Introduction

Our aim in this paper is to revisit the Breuil–Mézard conjecture [BM02] from a geometric point of view. Let us explain what we mean by this. First recall that the Breuil–Mézard conjecture posits a formula (in terms of certain representation-theoretic data) for the Hilbert–Samuel multiplicity of the characteristic p fibre of certain local $\mathbb{Z}_p$-algebras, namely those whose characteristic zero fibres parameterize two-dimensional potentially semistable liftings of some fixed continuous two-dimensional Galois representation $\bar{r} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\mathbb{F})$, where $\mathbb{F}$ is a finite field of characteristic p (the so-called *potentially semistable deformation rings* constructed in [Kis08]). One way in which a local ring can have multiplicity is if its Spec has more than one component: if its Spec is the union of n irreducible components, each with multiplicities μ_i $(i = 1, \ldots, n)$, then the multiplicity of the entire ring will be $\sum_i \mu_i$. Our goal is to both explain and refine the Breuil–Mézard conjecture in these terms, by identifying the irreducible components of the various rings involved, in representation-theoretic terms, as well as to determine their multiplicities.

To be somewhat more precise, after recalling some background material in Section 2, in Section 3 we consider the case of two-dimensional representations of $G_{\mathbb{Q}_p}$, as introduced above. In this case the Breuil–Mézard conjecture is a theorem of Kisin [Kis09a] (under very mild assumptions on $\bar{r}$), and we are able to strengthen Kisin's result so as to prove our geometric refinement of the conjecture. (We give a more detailed description of our results in this case in Subsection 1.1 below.) The possibility of such an extension is strongly suggested by the recent paper [BM12], and our results may be viewed as a sharpening of the results of *ibid.* (see Remark 1.1.9 below). In Section 4 we propose an extension of the Breuil–Mézard conjecture, and of our geometric refinement thereof, to the case of n-dimensional representations of G_K, for any finite extension K of $\mathbb{Q}_p$ and any positive integer n. Finally,

The first author was supported in part by NSF grant DMS-1003339, and the second author by NSF grant DMS-0841491.

in Section 5, we explain how the arguments of Section 3 may be extended to the case of n-dimensional representations, so as to prove an equivalence between the Breuil–Mézard conjecture (extended to the n-dimensional case) and its geometric refinement, under the assumption of a suitable $R = \mathbb{T}$-type theorem, together with a strong form of the weight part of Serre's conjecture for rank n unitary groups. In the case of two-dimensional potentially Barsotti–Tate representations, we deduce an unconditional geometric refinement of the results of [GK12].

In an appendix we establish a technical result that allows us to realize representations of local Galois groups as restrictions of automorphic representations of global Galois groups.

1.1. **Summary of our results in the case of two-dimensional representations of $G_{\mathbb{Q}_p}$.** We now explain in more detail our geometric refinement of the original Breuil–Mézard conjecture. To this end, we fix a finite extension E of $\mathbb{Q}_p$, with ring of integers $\mathcal{O}$, residue field $\mathbb{F}$, and uniformizer π. As above, we also fix a continuous representation $\bar{r} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\mathbb{F})$, and we let $R^{\square}(\bar{r})$ denote the universal lifting ring of $\bar{r}$ over $\mathcal{O}$. If m, n are integers with $n \geq 0$ and τ is an inertial type defined over E, then we may consider the subset of $\operatorname{Spec} R^{\square}(\bar{r})[1/p]$ consisting of those closed points that correspond to lifts of $\bar{r}$ to characteristic zero which are potentially semistable with Hodge–Tate weights $(m, m+n+1)$ and inertial type τ. (We adopt the convention that the cyclotomic character has Hodge–Tate weight 1, though we caution the reader that this convention does not remain in force for the entire paper; see Section 1.3 for the precise conventions we will follow.) In [Kis08], Kisin proves that there is a reduced closed subscheme $\operatorname{Spec} R^{\square}(m, n, \tau, \bar{r})$ of $\operatorname{Spec} R^{\square}(\bar{r})$ such that this subset is precisely the set of closed points of $\operatorname{Spec} R^{\square}(m, n, \tau, \bar{r})[1/p]$.[1] The Breuil–Mézard conjecture addresses the problem of describing the characteristic p fibre of $\operatorname{Spec} R^{\square}(m, n, \tau, \bar{r})$, i.e. the closed subscheme $\operatorname{Spec} R^{\square}(m, n, \tau, \bar{r})/\pi$ of $\operatorname{Spec} R^{\square}(\bar{r})/\pi$. More precisely, the conjecture as originally stated in [BM02] gives a conjectural formula for the Hilbert–Samuel multiplicity of this local scheme. This conjecture was proved (under very mild assumptions on $\bar{r}$) in [Kis09a]. In this paper we will prove a more precise statement, namely we will identify the underlying cycle of $\operatorname{Spec} R^{\square}(m, n, \tau, \bar{r})/\pi$; that is, we will describe the irreducible components of this scheme, and the multiplicity with which each component appears. To explain this more carefully, suppose first that $\mathcal{X}$ is any Noetherian scheme. If $\mathcal{Z}$ is a closed subscheme of $\mathcal{X}$, and $\mathfrak{p}$ is any point of $\mathcal{X}$, then we may define the (Hilbert–Samuel) multiplicity $e(\mathcal{Z}, \mathfrak{p})$ of $\mathcal{Z}$ at $\mathfrak{p}$ to be the Hilbert–Samuel multiplicity of the stalk $\mathcal{O}_{\mathcal{Z},\mathfrak{p}}$. Suppose now that $\mathcal{Z}$ is equidimensional of dimension d. If $\mathfrak{a}$ is a point of $\mathcal{X}$ of dimension d (i.e. whose closure $\overline{\{\mathfrak{a}\}}$ is of dimension d), then the stalk $\mathcal{O}_{\mathcal{Z},\mathfrak{a}}$ is either zero (if $\mathfrak{a} \notin \mathcal{Z}$) or an Artinian local ring (if $\mathfrak{a} \in \mathcal{Z}$, i.e. if $\mathfrak{a}$ is a generic point of $\mathcal{Z}$, or, equivalently, if $\overline{\{\mathfrak{a}\}}$ is an irreducible component of $\mathcal{Z}$), and the multiplicity $e(\mathcal{Z}, \mathfrak{a})$ is simply the length of $\mathcal{O}_{\mathcal{Z},\mathfrak{a}}$ as a module over itself, a quantity which can be interpreted geometrically as the multiplicity with which the component $\overline{\{\mathfrak{a}\}}$ appears in $\mathcal{Z}$. Since $\mathcal{Z}$ contains only finitely many generic points, the formal sum $Z(\mathcal{Z}) := \sum_{\mathfrak{a}} e(\mathcal{Z}, \mathfrak{a})\mathfrak{a}$ is well-defined as a d-dimensional cycle on $\mathcal{X}$, and we refer to

[1] In fact, in [Kis08] Kisin also fixes the determinants of the lifts that he considers, but we will suppress this technical point for now.

it as the cycle associated to $\mathcal{Z}$. If $\mathfrak{p}$ is any point of $\mathcal{X}$, then one has the formula

$$e(\mathcal{Z},\mathfrak{p}) = \sum_{\mathfrak{a}} e(\mathcal{Z},\mathfrak{a}) e\big(\big(\overline{\{\mathfrak{a}\}}\big),\mathfrak{p}\big) \tag{1.1.1}$$

(where again the sum ranges over points $\mathfrak{a}$ of dimension d), allowing us to compute the multiplicity of $\mathcal{Z}$ at any point in terms of its associated cycle. We are interested in the case when $\mathcal{X} := \operatorname{Spec} R^{\square}(\bar{r})/\pi$ (an 8-dimensional Noetherian local scheme) and $\mathcal{Z} := \operatorname{Spec} R^{\square}(m,n,\tau,\bar{r})/\pi$ for some m, n, τ. It is a theorem of [Kis08] that each of these closed subschemes $\mathcal{Z}$ is equidimensional of dimension 5, and so we may define the associated cycles $Z\big(\operatorname{Spec} R^{\square}(m,n,\tau,\bar{r})/\pi\big)$. Using this construction, we may in particular define a certain cycle on $\operatorname{Spec} R^{\square}(\bar{r})/\pi$ attached to each Serre weight of $\bar{r}$.

1.1.2. **Definition.** If σ is a Serre weight of $\bar{r}$, write $\sigma = \sigma_{m,n} := \det^m \otimes \operatorname{Sym}^n \mathbb{F}^2$ for integers m, n such that $0 \le n \le p-1$, and define $\mathcal{C}_{m,n} := Z\big(\operatorname{Spec} R^{\square}(m,n,\mathbb{1},\bar{r})/\pi\big)$. (To avoid ambiguity, one could insist that m is chosen so that $0 \le m \le p-2$. However, the subscheme $\operatorname{Spec} R^{\square}(m,n,\mathbb{1},\bar{r})/\pi$ is in fact independent of the particular choice of m used to describe σ.)

The following proposition describes these cycles quite explicitly. We will say that $\mathcal{C}_{m,n}$ consists of a single component if it has a single irreducible component, which is also reduced, and that it consists of two components if it consists of two reduced and irreducible components.

1.1.3. **Proposition.** *Assume that the Breuil–Mézard Conjecture* (*i.e. Conjecture* 3.1.1 *below*) *holds for* $\bar{r}$.

(1) *If $\bar{r}$ is irreducible and $\sigma_{m,n}$ is a Serre weight of $\bar{r}$, then $\mathcal{C}_{m,n}$ consists of a single component, which has multiplicity one at the closed point of* $\operatorname{Spec} R^{\square}(\bar{r})/\pi$.
(2) *If $\bar{r}$ is reducible and $\sigma_{m,n}$ is a Serre weight of $\bar{r}$ such that $n < p-2$, or $n = p-2$ and $\bar{r}$ is a non-split extension of distinct characters, then $\mathcal{C}_{m,n}$ consists of a single component, which has multiplicity one at the closed point of* $\operatorname{Spec} R^{\square}(\bar{r})/\pi$.
(3) *If $\bar{r}$ is reducible and $\sigma_{m,n}$ is a Serre weight of $\bar{r}$ with $n = p-1$, so that $\bar{r}|_{I_p} \sim \begin{pmatrix} \omega^{m+1} & * \\ 0 & \omega^m \end{pmatrix}$, then if $*$ is peu ramifiée and $\bar{r}$ itself is a twist of an extension of the trivial character by the mod p cyclotomic character, then $\mathcal{C}_{m,n}$ is a sum of two components, each having multiplicity one at the closed point of* $\operatorname{Spec} R^{\square}(\bar{r})/\pi$. *Otherwise $\mathcal{C}_{m,n}$ is a single component, having multiplicity one at the closed point of* $\operatorname{Spec} R^{\square}(\bar{r})/\pi$.
(4) *If $\sigma_{m,n}$ is a Serre weight of $\bar{r}$ with $n = p-2$ and $\bar{r}$ is split and p-distinguished, then $\mathcal{C}_{m,n}$ is a sum of two components, each having multiplicity one at the closed point of* $\operatorname{Spec} R^{\square}(\bar{r})/\pi$.
(5) *If $\sigma_{m,n}$ is a Serre weight of $\bar{r}$ with $n = p-2$ and $\bar{r}$ has scalar semisimplification, then $\mathcal{C}_{m,n}$ consists of a single component.*
(6) *If $\sigma_{m,n}$ and $\sigma_{m',n'}$ are distinct Serre weights of $\bar{r}$, then $\mathcal{C}_{m,n}$ and $\mathcal{C}_{m',n'}$ have disjoint support, except if $m \equiv m' \pmod{p-1}$, $n = 0$ and $n' = p-1$ (possibly after interchanging $\sigma_{m,n}$ and $\sigma_{m',n'}$), in which case $\mathcal{C}_{m,n}$ is equal to a component of $\mathcal{C}_{m',n'}$.*

We remark that by the results of [Kis09a], the hypothesis of the preceeding proposition holds for most $\bar{r}$.

We make one more definition before stating our main theorem.

1.1.4. **Definition.** Given integers a, b with $b \geq 0$ and an inertial type τ (assumed to be defined over E), let $\sigma(\tau)$ denote the representation of $\mathrm{GL}_2(\mathbb{Z}_p)$ over E associated to τ via Henniart's inertial local Langlands correspondence, write $\sigma(a,b,\tau) := (\det^a \otimes \mathrm{Sym}^b E^2) \otimes_E \sigma(\tau)$, and let $\overline{\sigma(a,b,\tau)}^{\mathrm{ss}}$ denote the semi-simplification of the reduction mod π of (any) $\mathrm{GL}_2(\mathbb{Z}_p)$-invariant $\mathcal{O}$-lattice in $\sigma(a,b,\tau)$. (The representation so obtained is well-defined independent of the choice of invariant lattice.)

We may now state our geometric refinement of the Breuil–Mézard conjecture.

1.1.5. **Theorem.** *Suppose that*

$$\bar{r} \not\sim \begin{pmatrix} \omega\chi & * \\ 0 & \chi \end{pmatrix}$$

for any character χ. Fix integers m, n with $n \geq 0$ and an inertial type τ, and for each Serre weight $\sigma_{m,n}$ of $\bar{r}$, let $a_{m,n}$ denote the multiplicity with which $\sigma_{m,n}$ appears as a constituent of $\overline{\sigma(a,b,\tau)}^{\mathrm{ss}}$. Then we have the following equality of cycles:

$$Z\big(\mathrm{Spec}\, R^{\square}(a,b,\tau,\bar{r})/\pi\big) = \sum_{m,n} a_{m,n} \mathcal{C}_{m,n}.$$

1.1.6. **Remark.** The usual form of the Breuil–Mézard conjecture, as stated in [BM02] and proved in [Kis09a], can be recovered from this result by applying the formula (1.1.1), and using the explicit description of the cycles $\mathcal{C}_{m,n}$ provided by Proposition 1.1.3. (Note that while Proposition 1.1.3 as stated assumes that the Breuil–Mézard conjecture holds for $\bar{r}$, our proof of Proposition 1.1.3 will actually use Theorem 1.1.5 for $\bar{r}$ as input, and so this argument is not circular.)

1.1.7. **Remark.** The hypothesis on $\bar{r}$ in Theorem 1.1.5 is slightly weaker than that made in the analogous result in [Kis09a]. This is due to our use of potential modularity theorems to realise local representations globally, which is more flexible than the construction of [Kis09a] using CM forms. We remark that Paškūnas ([Paš12]) has reproved Kisin's results (and our generalisation of them) by purely local means under a similarly weakened hypothesis on $\bar{r}$.

1.1.8. **Remark.** Theorem 1.1.5 is proved via a refinement of the global argument made in [Kis09a], and uses the local arguments of [Kis09a] (using the p-adic Langlands correspondence) as an input. In particular, it does not give a new proof of the usual form of the Breuil–Mézard conjecture.

1.1.9. **Remark.** In the recent sequel [BM12] to their paper [BM02], Breuil and Mézard have constructed, for generic $\bar{r}$, a correspondence between the irreducible components of $\mathrm{Spec}\, R^{\square}(a,b,\tau,\bar{r})/\pi$ and the Serre weights. We show (in Subsection 3.4) that this coincides with the correspondence $\sigma_{m,n} \mapsto \mathcal{C}_{m,n}$. (Note that when $\bar{r}$ is generic in the sense of [BM12], each cycle $\mathcal{C}_{m,n}$ consists of a single component.) Thus our results may be reviewed as a refinement of those of [BM12]. We also note that in [BM12], the authors proceed by refining the local arguments of [Kis09a], while (as already noted) in this note we proceed by refining the global arguments of *ibid.* Thus the approaches of [BM12] and of the present note may be regarded as being somewhat complementary to one another.

1.1.10. **Remark.** In the paper [GK12], similar techniques to those of this paper are used to prove the Breuil–Mézard conjecture for two-dimensional potentially Barsotti–Tate representations of G_K. The key additional ingredients which are available in that case, but not in general, are the automorphy lifting theorems for potentially Barsotti–Tate representations proved in [Kis09b] and [Gee06]. The arguments of the present paper are in large part based on those of [GK12], which in turn relies on the strategy outlined in [Kis10]; in particular, our implementation of the patching argument for unitary groups is simply the natural adaptation of the arguments of [GK12] to higher rank unitary groups. Theorem 5.5.4 below gives a geometric refinement of some of the results of [GK12].

1.2. **Acknowledgments.** The debt that the arguments of this paper owe to the work of Mark Kisin will be obvious to the reader; in particular, several of our main arguments are closely based on arguments from [Kis09a], [Kis10] and [GK12]. We are also grateful to him, as well as to Christophe Breuil and Kevin Buzzard, for helpful comments on an earlier draft of this paper. We would like to thank Florian Herzig for helpful conversations about the representation theory of $\mathrm{GL}_n(\mathbb{F}_q)$, and Tom Barnet-Lamb, David Geraghty, Robert Guralnick, Florian Herzig, and Jack Shotton for helpful conversations about the material in Appendix A. We would like to thank the anonymous referee for a careful reading, and many helpful comments and suggestions.

1.3. **Notation and Conventions.** Throughout this paper, p will denote an odd prime.

If K is a field, then we let G_K denote its absolute Galois group. If K is furthermore a finite extension of $\mathbb{Q}_p$ for some p, then we write I_K for the inertia subgroup of G_K. 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}.

We let ε denote the p-adic cyclotomic character, and let $\overline{\varepsilon} = \omega$ the mod p cyclotomic character. We denote by $\widetilde{\omega}$ the Teichmüller lift of ω. We let ω_2 denote a choice of a fundamental character of $I_{\mathbb{Q}_p}$ of niveau 2.

If K is a p-adic field, if ρ is a continuous de Rham representation of G_K over $\overline{\mathbb{Q}}_p$, and if $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$, then we will write $\mathrm{HT}_\tau(\rho)$ for the multiset of Hodge–Tate numbers of ρ with respect to τ. By definition, if W is a de Rham representation of G_K over $\overline{\mathbb{Q}}_p$ and if $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$ then the multiset $\mathrm{HT}_\tau(W)$ contains i with multiplicity $\dim_{\overline{\mathbb{Q}}_p}(W \otimes_{\tau,K} \widehat{\overline{K}}(i))^{G_K}$. Thus for example $\mathrm{HT}_\tau(\varepsilon) = \{-1\}$. We will use this convention throughout the paper, except in Section 3, where we will use the opposite convention that ε has Hodge–Tate weight 1. We apologise for this, but it seems to us to be the best way to make what we write compatible with the existing literature.

Let K be a finite extension of $\mathbb{Q}_p$, and let rec denote the local Langlands correspondence from isomorphism classes of irreducible smooth representations of $\mathrm{GL}_n(K)$ over $\mathbb{C}$ to isomorphism classes of n-dimensional Frobenius semisimple Weil–Deligne representations of W_K defined in [HT01]. Fix an isomorphism $\imath : \overline{\mathbb{Q}}_p \to \mathbb{C}$. We define the local Langlands correspondence rec_p over $\overline{\mathbb{Q}}_p$ by $\imath \circ \mathrm{rec}_p = \mathrm{rec} \circ \imath$. This depends only on $\imath^{-1}(\sqrt{p})$. We let $\mathrm{Art}_K : K^\times \to W_K^{ab}$ be the isomorphism provided by local class field theory, which we normalise so that uniformisers correspond to geometric Frobenius elements. We will write $\mathbb{1}$ for the trivial n-dimensional representation of some group, the precise group and choice of n always being clear from the context.

When discussing deformations of Galois representations, we will use the terms "framed deformation" (which originated in [Kis09b]) and "lifting" (which originated in [CHT08]) interchangeably.

We write all matrix transposes on the left; so ${}^tA$ is the transpose of A.

2. Background on multiplicities and cycles

In this preliminary section we provide details on the notions of multiplicities and cycles that we outlined in the introduction.

2.1. **Hilbert–Samuel multiplicities.** Recall that if A is a Noetherian local ring with maximal ideal $\mathfrak{m}$ of dimension d, and M is a finite A-module, then there is polynomial $P_M^A(X)$ of degree at most d (the Hilbert–Samuel polynomial of M), uniquely determined by the requirement that for $n \gg 0$, the value $P_M^A(n)$ is equal to the length of $M/\mathfrak{m}^{n+1}M$ as an A-module.

2.1.1. **Definition.** The Hilbert–Samuel multiplicity $e(M, A)$ is defined to be $d!$ times the coefficient of X^d in $P_M^A(X)$. We write $e(A)$ for $e(A, A)$.

Note in particular that if A is Artinian, then $e(M, A)$ is simply the length of M as an A-module.

2.2. **Cycles.** Let $\mathcal{X}$ be a Noetherian scheme.

2.2.1. **Definition.**
(1) Let $\mathcal{M}$ be a coherent sheaf on $\mathcal{X}$, and write $\mathcal{Z}$ to denote the scheme-theoretic support of $\mathcal{M}$ (i.e. $\mathcal{Z}$ is the closed subscheme of $\mathcal{X}$ cut out by the annihilator ideal $\mathcal{I} \subset \mathcal{O}_X$ of $\mathcal{M}$). For any point $x \in \mathcal{X}$, we write $e(\mathcal{M}, x)$ to denote the Hilbert–Samuel multiplicity $e(\mathcal{M}_x, \mathcal{O}_{\mathcal{Z},x})$.
(2) If $\mathcal{Z}$ is a closed subscheme of $\mathcal{X}$, then we write $e(\mathcal{Z}, x) := e(\mathcal{O}_{\mathcal{Z}}, x)$ for all $x \in X$. (Note that $\mathcal{Z}$ coincides with the scheme-theoretic support of $\mathcal{O}_{\mathcal{Z}}$, and so by definition this is equal to the multiplicity $e(\mathcal{O}_{\mathcal{Z},x})$ of the local ring $\mathcal{O}_{\mathcal{Z},x}$.)

2.2.2. **Remark.** In some situations the multiplicity $e(\mathcal{M}, x)$ is particularly simple to describe.

(1) If x does not lie in the support of $\mathcal{M}$, i.e. if $\mathcal{M}_x = 0$, then $e(\mathcal{M}, x) = 0$.
(2) If x is a generic point of the scheme-theoretic support $\mathcal{Z}$ of $\mathcal{M}$ (so that $\mathcal{O}_{\mathcal{Z},x}$ is an Artinian ring), then $e(\mathcal{M}, x)$ is simply the length of $\mathcal{M}_x$ as an $\mathcal{O}_{\mathcal{Z},x}$-module.

2.2.3. **Definition.**
(1) We say that a point $x \in X$ is of dimension d, and write $\dim(x) = d$, if its closure $\overline{\{x\}}$ is of dimension d.
(2) A d-dimensional cycle on $\mathcal{X}$ is a formal finite $\mathbb{Z}$-linear combination of points of $\mathcal{X}$ of dimension d.
(3) We write $\mathcal{X} \geq 0$ if $\mathcal{X}$ is in fact a $\mathbb{Z}_{\geq 0}$-linear combination of points of $\mathcal{X}$ of dimension d, and we write $\mathcal{X} \geq \mathcal{Y}$ if $\mathcal{X} - \mathcal{Y} \geq 0$.

2.2.4. **Definition.** If $Z = \sum_{\dim(x)=d} n_x x$ is a d-dimensional cycle on $\mathcal{X}$, then for any point $y \in \mathcal{X}$, we define the multiplicity $e(Z, y)$ via the formula

$$e(Z, y) := \sum_{\dim(x)=d} n_x e(\overline{\{x\}}, y).$$

(Here $\overline{\{x\}}$ denotes the closure of the point x.)

2.2.5. **Definition.** (1) If $d \geq 0$ is a non-negative integer, and $\mathcal{M}$ is a coherent sheaf on $\mathcal{X}$ whose support has dimension $\leq d$, then we define the d-dimensional cycle $Z_d(\mathcal{M})$ associated to $\mathcal{M}$ as follows:

$$Z_d(\mathcal{M}) := \sum_{\dim(x)=d} e(\mathcal{M}, x)x,$$

where, as indicated, the sum ranges over all points of $\mathcal{X}$ of dimension d. (Our assumption on the dimension of the support of $\mathcal{M}$ ensures that any point of dimension d lying in the support $\mathcal{M}$ is necessarily a generic point of that support, and hence that there are only finitely many such points lying in the support of $\mathcal{M}$. Thus all but finitely many terms appearing in the sum defining $Z_d(\mathcal{M})$ vanish, and so this sum is in fact well-defined. Note also that, if we let $\mathcal{Z}$ denote the scheme-theoretic support of $\mathcal{M}$, then by Remark 2.2.2, the multiplicity $e(\mathcal{M}, x)$ is simply the length of $\mathcal{M}_x$ as an $\mathcal{O}_{\mathcal{Z},x}$-module.)

(2) If the support of $\mathcal{M}$ is finite-dimensional of dimension d, then we write simply $Z(\mathcal{M}) := Z_d(\mathcal{M})$.

(3) If $\mathcal{Z}$ is a closed subset of $\mathcal{X}$, then we write $Z_d(\mathcal{Z}) := Z_d(\mathcal{O}_{\mathcal{Z}})$, and denote this simply by $Z(\mathcal{Z})$ if $\mathcal{Z}$ is finite-dimensional of dimension d.

2.2.6. **Remark.** If $\mathcal{Z}$ is equidimensional of some finite dimension d, then $Z(\mathcal{Z})$ encodes the irreducible components of $\mathcal{Z}$, together with the multiplicity with which each component appears in $\mathcal{Z}$. If $\mathcal{Z}$ has dimension less than d, then $Z_d(\mathcal{Z}) = 0$.

2.2.7. **Lemma.** *If $0 \to \mathcal{M}' \to \mathcal{M} \to \mathcal{M}'' \to 0$ is an exact sequence of coherent sheaves on $\mathcal{X}$, such that the support of $\mathcal{M}$ is of dimension $\leq d$ (or equivalently, such that the supports of each of $\mathcal{M}'$ and $\mathcal{M}''$ are of dimension $\leq d$), then*

$$Z_d(\mathcal{M}) = Z_d(\mathcal{M}') + Z_d(\mathcal{M}'').$$

Proof. Let $\mathcal{Z}$ (resp. $\mathcal{Z}'$ and $\mathcal{Z}''$) denote the support of $\mathcal{M}$ (resp. of $\mathcal{M}'$ and $\mathcal{M}''$), so that $\mathcal{Z}', \mathcal{Z}'' \subseteq \mathcal{Z}$. As already noted in the statement of Definition 2.2.5, if $x \in \mathcal{X}$ is of dimension d, then $e(\mathcal{M}, x)$ (resp. $e(\mathcal{M}', x)$, resp. $e(\mathcal{M}'', x)$) is simply the length of $\mathcal{M}_x$ as an $\mathcal{O}_{\mathcal{Z},x}$-module (resp. the length of $\mathcal{M}'_x$ as an $\mathcal{O}_{\mathcal{Z}',x}$-module, resp. the length of $\mathcal{M}''_x$ as an $\mathcal{O}_{\mathcal{Z}'',x}$-module). Since each of $\mathcal{O}_{\mathcal{Z}',x}$ and $\mathcal{O}_{\mathcal{Z}'',x}$ is a quotient of $\mathcal{O}_{\mathcal{Z},x}$, the claimed additivity of cycles follows from the additivity of lengths in exact sequences. □

The following lemma records formula (1.1.1), which was stated in the introduction.

2.2.8. **Lemma.** *If $\mathcal{M}$ is a coherent sheaf on $\mathcal{X}$ with support of dimension d, and if $Z(\mathcal{M})$ is the cycle associated to $\mathcal{M}$, then for any point $y \in X$, we have the formula*

$$e(\mathcal{M}, y) = e\big(Z(\mathcal{M}), y\big).$$

Proof. This follows immediately from Theorem 14.7 of [Mat89]. □

We now establish a cycle-theoretic version of [Kis09a, Prop. 1.3.4], whose proof follows the same lines as the proof of that result.

2.2.9. **Lemma.** *Let $\mathcal{X}$ be a Noetherian scheme of finite dimension d, and let $f \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ be regular (i.e. a non-zero divisor in each stalk of $\mathcal{O}_{\mathcal{X}}$). If $\mathcal{M}$ is an f-torsion free coherent sheaf on $\mathcal{X}$ which is supported in dimension $d-1$, then $\mathcal{M}/f\mathcal{M}$ is supported in dimension $d-2$.*

Proof. Since $\mathcal{M}$ is f-torsion free, no generic point of $\mathrm{Supp}(\mathcal{M})$ is contained in $V(f)$. Thus $\mathrm{Supp}(\mathcal{M}/f\mathcal{M})$ is of dimension $\leq d-2$, as claimed. □

2.2.10. **Lemma.** *Let $\mathcal{X}$ be a Noetherian integral scheme of finite dimension d, and let $f \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ be non-zero. If $\mathcal{M}$ is an f-torsion free coherent sheaf on $\mathcal{X}$ which is generically free of rank one, then $Z_{d-1}(\mathcal{M}/f\mathcal{M}) = Z\big(V(f)\big)$.*

Proof. Let x be the generic point of $\mathcal{X}$, and $i_x : \operatorname{Spec} \kappa(x) \to \mathcal{X}$ the canonical map. By assumption $\mathcal{M}_x$ is one-dimensional over $\kappa(x)$, and so we may find a morphism of quasi-coherent sheaves $\mathcal{M} \to (i_x)_*\kappa(x)$ whose kernel $\mathcal{M}'$ is torsion. The image $\mathcal{M}''$ of this morphism is a coherent subsheaf of $(i_x)_*\kappa(x)$, and hence is contained in an invertible sheaf $\mathcal{L}$. Let $\mathcal{L}'$ denote the cokernel of the inclusion $\mathcal{M}'' \hookrightarrow \mathcal{L}$.

Consider first the exact sequence

$$0 \to \mathcal{M}' \to \mathcal{M} \to \mathcal{M}'' \to 0.$$

All the terms in this sequence are f-torsion free, and so

$$0 \to \mathcal{M}'/f\mathcal{M}' \to \mathcal{M}/f\mathcal{M} \to \mathcal{M}''/f\mathcal{M}'' \to 0$$

is again exact. Lemma 2.2.9 shows that $Z_{d-1}(\mathcal{M}'/f\mathcal{M}') = 0$, and Lemma 2.2.7 then implies that

$$Z_{d-1}(\mathcal{M}/f\mathcal{M}) = Z_{d-1}(\mathcal{M}''/f\mathcal{M}''). \tag{2.2.11}$$

Next consider the exact sequence $0 \to \mathcal{L}'[f] \to \mathcal{L}' \to \mathcal{L}' \to \mathcal{L}'/f\mathcal{L}' \to 0$. Since all the terms in this exact sequence are torsion, and so supported in dimension $d-1$, we see from Lemma 2.2.7 that

$$Z_{d-1}(\mathcal{L}'[f]) = Z_{d-1}(\mathcal{L}'/f\mathcal{L}'). \tag{2.2.12}$$

Finally, consider the exact sequence $0 \to \mathcal{M}'' \to \mathcal{L} \to \mathcal{L}' \to 0$. The first two terms in this sequence are f-torsion free, and so it induces an exact sequence $0 \to \mathcal{L}'[f] \to \mathcal{M}''/f\mathcal{M}'' \to \mathcal{L}/f\mathcal{L} \to \mathcal{L}'/f\mathcal{L}' \to 0$. Again applying Lemma 2.2.7, together with (2.2.12), we find that $Z_{d-1}(\mathcal{M}''/f\mathcal{M}'') = Z_{d-1}(\mathcal{L}/f\mathcal{L})$. Combining this with (2.2.11), together with the fact that $\mathcal{L}$ is an invertible sheaf, we find that $Z_{d-1}(\mathcal{M}/f\mathcal{M}) = Z\big(V(f)\big)$, as required. □

2.2.13. **Proposition.** *Let $\mathcal{X}$ be a Noetherian scheme of finite dimension d, and $f \in \mathcal{O}_{\mathcal{X}}(\mathcal{X})$ be regular (i.e. a non-zero divisor in each stalk of $\mathcal{O}_{\mathcal{X}}$). If $\mathcal{M}$ is an f-torsion free coherent sheaf on $\mathcal{X}$, and if $Z_d(\mathcal{M}) = \sum_{\dim(x)=d} n_x x$, where, as indicated, x runs over the d-dimensional points of $\mathcal{X}$, then the support of $\mathcal{M}/f\mathcal{M}$ has dimension $\leq d-1$, and*

$$Z_{d-1}(\mathcal{M}/f\mathcal{M}) = \sum_{\dim(x)=d} n_x Z_{d-1}\big(\overline{\{x\}} \cap V(f)\big).$$

Proof. We argue by induction on the quantity $n := \sum_{\dim x=d} n_x$, with the case when $n = 0$ being handled by Lemma 2.2.9. Suppose now that n is an arbitrary positive integer, and choose x of dimension d such that $n_x > 0$. We may then find a non-zero surjection $\mathcal{M}_x \to \kappa(x)$ (since by definition n_x is the length of $\mathcal{M}_x$), which induces a non-zero map of quasi-coherent sheaves $\mathcal{M} \to (i_x)_*\kappa(x)$ (where i_x is the canonical map $\operatorname{Spec}\kappa(x) \to \mathcal{X}$). Let $\mathcal{M}'$ denote the kernel of this map, and $\mathcal{M}''$ its image, so that the sequence $0 \to \mathcal{M}' \to \mathcal{M} \to \mathcal{M}'' \to 0$ is exact.

The sheaf $\mathcal{M}'$ is f-torsion free (since it is a subsheaf of $\mathcal{M}$) while the sheaf $\mathcal{M}''$ is f-torsion free, supported on $\overline{\{x\}}$, and generically free of rank one over this component (being a non-zero coherent subsheaf of $(i_x)_*\kappa(x)$). Thus we obtain a short exact sequence

$$0 \to \mathcal{M}'/f\mathcal{M}' \to \mathcal{M}/f\mathcal{M} \to \mathcal{M}''/f\mathcal{M}'' \to 0,$$

and Lemmas 2.2.7 and 2.2.10 show that

$$\begin{aligned} Z_{d-1}(\mathcal{M}/f\mathcal{M}) = Z_{d-1}(\mathcal{M}'/f\mathcal{M}') + Z_{d-1}(\mathcal{M}''/f\mathcal{M}'') \\ = Z_{d-1}(\mathcal{M}'/f\mathcal{M}') + Z_{d-1}\big(\overline{\{x\}} \cap V(f)\big). \end{aligned}$$

The proposition follows by induction. □

We close this section with a result about the product of cycles. In fact, we will need to apply such a result in the context of a completed tensor product of complete Noetherian local k-algebras, for some field k (which is fixed for the remainder of this discussion), and so we restrict our attention to that particular context.

Note that if A and B are complete Noetherian local k-algebras, and if $\mathfrak{p}$ and $\mathfrak{q}$ are primes of A and B respectively, such that $A/\mathfrak{p}$ is of dimension d and $B/\mathfrak{q}$ is of dimension e, then $A/\mathfrak{p}\widehat{\otimes}_k B/\mathfrak{q}$ is a quotient of $A\widehat{\otimes}_k B$ of dimension $d+e$. Hence $\operatorname{Spec} A/\mathfrak{p}\widehat{\otimes}_k B/\mathfrak{q}$ is a closed subscheme of $\operatorname{Spec} A\widehat{\otimes}_k B$ of dimension $d+e$, and we write

$$Z(\operatorname{Spec} A/\mathfrak{p}) \times_k Z(\operatorname{Spec} B/\mathfrak{q}) := Z(\operatorname{Spec} A/\mathfrak{p}\widehat{\otimes}_k B/\mathfrak{q}),$$

and then extend this by linearity to a bilinear product from d-dimensional cycles on $\operatorname{Spec} A$ and e-dimensional cycles on $\operatorname{Spec} B$ to $(d+e)$-dimensional cycles on $\operatorname{Spec} A\widehat{\otimes}_k B$.

If M and N are finitely generated A- and B-modules respectively, giving rise to coherent sheaves $\mathcal{M}$ and $\mathcal{N}$ on $\operatorname{Spec} A$ and $\operatorname{Spec} B$ respectively, then the completed tensor product $M\widehat{\otimes}_k N$ gives rise to a coherent sheaf on $\operatorname{Spec} A\widehat{\otimes}_k B$, which we denote by $\mathcal{M}\widehat{\boxtimes}\mathcal{N}$.

2.2.14. **Lemma.** *In the context of the preceding discussion, if the support of $\mathcal{M}$ and $\mathcal{N}$ are of dimensions d and e respectively, then the support of $\mathcal{M}\widehat{\boxtimes}\mathcal{N}$ is of dimension $d+e$, and*

$$Z_{d+e}(\mathcal{M}\widehat{\boxtimes}\mathcal{N}) = Z_d(\mathcal{M}) \times_k Z_e(\mathcal{N}).$$

Proof. This is standard; we sketch the proof. Restricting to the support of $\mathcal{M}\widehat{\boxtimes}\mathcal{N}$, we may assume that M (resp. N) is a faithful A-module (resp. B-module), so that A has dimension d and B has dimension e. As in (for example) the proof of [BLGHT11, Lemma 3.3(5)], the distinct minimal primes of $A\widehat{\otimes}_k B$ are precisely the $\mathfrak{p}(A\widehat{\otimes}_k B)+\mathfrak{q}(A\widehat{\otimes}_k B)$, where $\mathfrak{p}$ is a minimal prime of A and $\mathfrak{q}$ is a minimal prime of B. We are thus reduced to checking that for each such $\mathfrak{p}, \mathfrak{q}$, we have

$$e(M\widehat{\otimes}_k N, (A/\mathfrak{p})\widehat{\otimes}_k(B/\mathfrak{q})) = e(M, (A/\mathfrak{p}))e(N, (B/\mathfrak{q})),$$

which follows from Lech's lemma [Mat89, Thm. 14.12] exactly as in the proof of Proposition 1.3.8 of [Kis09a]. □

3. The geometric Breuil–Mézard Conjecture for two-dimensional representations of $G_{\mathbb{Q}_p}$

In this section we explain the Breuil–Mézard conjecture in its original setting, that of two-dimensional representations of $G_{\mathbb{Q}_p}$, and state our geometric version. We then recall some of the details of the proof of the original formulation of the conjecture from [Kis09a], and show that the proof may be extended to prove the geometric version. The one difference from Kisin's notation and that of the present paper is that we prefer not to fix one of the Hodge–Tate weights of our Galois representations to be 0; this makes no essential difference to any of the arguments, but the additional flexibility that this notation gives us is convenient in the exposition. We remind the reader that in this section, our conventions for Hodge–Tate weights are that ε has Hodge–Tate weight 1.

3.1. **The conjecture.** We begin by recalling some notation from [Kis09a]. Fix a prime $p > 2$ and a finite extension E of $\mathbb{Q}_p$ (our coefficient field), with ring of integers $\mathcal{O}$, residue field $\mathbb{F}$, and uniformiser π. We assume that $\#\mathbb{F} > 5$, so that $\mathrm{PSL}_2(\mathbb{F})$ is a simple group.

Let $\bar{r} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation, and let $\tau : I_{\mathbb{Q}_p} \to \mathrm{GL}_2(E)$ be an inertial type, i.e. a representation with open kernel which extends to $W_{\mathbb{Q}_p}$. Fix integers a, b with $b \geq 0$ and a de Rham character $\psi : G_{\mathbb{Q}_p} \to \mathcal{O}^\times$ such that $\overline{\psi\varepsilon} = \det \bar{r}$. We let $R^{\square,\psi}(a, b, \tau, \bar{r})$ and $R^{\square,\psi}_{\mathrm{cr}}(a, b, \tau, \bar{r})$ be the framed deformation $\mathcal{O}$-algebras which are universal for framed deformations of $\bar{r}$ which have determinant $\psi\varepsilon$, and are potentially semistable (respectively potentially crystalline) with Hodge–Tate weights $(a, a + b + 1)$ and inertial type τ. As in Section 1.1.2 of [Kis09a], we let $\sigma(\tau)$ and $\sigma_{\mathrm{cr}}(\tau)$ denote the finite-dimensional irreducible E-representations of $\mathrm{GL}_2(\mathbb{Z}_p)$ corresponding to τ via Henniart's inertial local Langlands correspondence, we set $\sigma(a, b, \tau) = (\det^a \otimes \mathrm{Sym}^b E^2) \otimes_E \sigma(\tau)$ and $\sigma^{\mathrm{cr}}(a, b, \tau) = (\det^a \otimes \mathrm{Sym}^b E^2) \otimes_E \sigma^{\mathrm{cr}}(\tau)$, and we let $L_{a,b,\tau}$ (respectively $L^{\mathrm{cr}}_{a,b,\tau}$) be a $\mathrm{GL}_2(\mathbb{Z}_p)$-stable $\mathcal{O}$-lattice in $\sigma(a, b, \tau)$ (respectively $\sigma^{\mathrm{cr}}(a, b, \tau)$). Write $\sigma_{m,n}$ for the representation $\det^m \otimes \mathrm{Sym}^n \mathbb{F}^2$ of $\mathrm{GL}_2(\mathbb{F}_p)$, $0 \leq m \leq p-2$, $0 \leq n \leq p-1$, so that we may write

$$(L_{a,b,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_{m,n} \sigma_{m,n}^{a_{m,n}},$$

and

$$(L^{\mathrm{cr}}_{a,b,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_{m,n} \sigma_{m,n}^{a^{\mathrm{cr}}_{m,n}},$$

for some integers $a_{m,n}$, $a^{\mathrm{cr}}_{m,n}$.

We can now state the Breuil–Mézard conjecture [BM02].

3.1.1. **Conjecture.** *There are integers $\mu_{m,n}(\bar{r})$ depending only on m, n, and $\bar{r}$, such that for any a, b, τ with $\det \tau = \varepsilon^{1-2a-b}\psi|_{I_{\mathbb{Q}_p}}$,*

$$e(R^{\square,\psi}(a, b, \tau, \bar{r})/\pi) = \sum_{m,n} a_{m,n}\mu_{m,n}(\bar{r}),$$

and

$$e(R^{\square,\psi}_{\mathrm{cr}}(a, b, \tau, \bar{r})/\pi) = \sum_{m,n} a^{\mathrm{cr}}_{m,n}\mu_{m,n}(\bar{r}).$$

3.1.2. **Remark.** By a straightforward twisting argument, the truth of the conjecture is independent of the choice of ψ, so we may assume that ψ is crystalline. Note that if the conjecture is true for all a, b, τ, then it is easy to see that $\mu_{m,n}(\bar{r}) = 0$

unless $\det \bar{r}|_{I_{\mathbb{Q}_p}} = \omega^{2m+n+1}$, and that if this holds then we must have $\mu_{m,n}(\bar{r}) = e(R_{\mathrm{cr}}^{\square,\psi}(\widetilde{m},n,\mathbb{1},\bar{r})/\pi)$, where $\widetilde{m}$ is chosen so that $\psi|_{I_{\mathbb{Q}_p}} = \varepsilon^{2\widetilde{m}+n+1}$. So all the values $\mu_{m,n}(\bar{r})$ are determined by the crystalline deformation rings in low weight.

3.1.3. **Remark.** Conjecture 3.1.1 was proved in [Kis09a] under the additional assumptions that $\bar{r} \not\sim \begin{pmatrix} \omega\chi & * \\ 0 & \chi \end{pmatrix}$ for any χ, and that if $\bar{r}$ has scalar semisimplification then $\bar{r}$ is scalar.

We may now state our geometric version of the Breuil–Mézard conjecture.

3.1.4. **Conjecture.** *For each $0 \le m \le p-2$, $0 \le n \le p-1$, there is a cycle $\mathcal{C}_{m,n}$ depending only on m, n, and $\bar{r}$, such that for any a, b, τ with* $\det \tau = \varepsilon^{-2a-b-1}\psi|_{I_{\mathbb{Q}_p}}$,

$$Z(R^{\square,\psi}(a,b,\tau,\bar{r})/\pi) = \sum_{m,n} a_{m,n}\mathcal{C}_{m,n},$$

and

$$Z(R_{\mathrm{cr}}^{\square,\psi}(a,b,\tau,\bar{r})/\pi) = \sum_{m,n} a_{m,n}^{\mathrm{cr}}\mathcal{C}_{m,n}.$$

3.1.5. **Remark.** Note again that if the conjecture is true for all choices of a, b, τ then with the assumptions and notation of Remark 3.1.2 we must have $\mathcal{C}_{m,n} = Z(R_{\mathrm{cr}}^{\square,\psi}(\widetilde{m},n,\mathbb{1},\bar{r})/\pi)$.

The following is our main result towards Conjecture 3.1.4.

3.1.6. **Theorem.** *If $\bar{r} \not\sim \begin{pmatrix} \omega\chi & * \\ 0 & \chi \end{pmatrix}$ for any χ, then Conjecture* 3.1.4 *holds for $\bar{r}$.*

3.2. **The proof of Theorem 3.1.6 via patching.** In this subsection we present the proof of Theorem 3.1.6. To this end, we fix a continuous representation $\bar{r} : G_{\mathbb{Q}_p} \to \mathrm{GL}_2(\mathbb{F})$. We begin by realising this representation as the restriction of a global Galois representation, by a similar (but simpler) argument to that of Appendix A of [GK12].

3.2.1. **Proposition.** *There is a totally real field F and a continuous irreducible representation $\overline{\rho} : G_F \to \mathrm{GL}_2(\mathbb{F})$ such that*

(1) *p splits completely in F;*
(2) *$\overline{\rho}$ is totally odd;*
(3) *$\overline{\rho}(G_F) = \mathrm{GL}_2(\mathbb{F})$;*
(4) *if $v \nmid p$ is a place of F then $\overline{\rho}|_{G_{F_v}}$ is unramified;*
(5) *if $v|p$ is a place of F then $\overline{\rho}|_{G_{F_v}} \cong \bar{r}$;*
(6) *$[F:\mathbb{Q}]$ is even;*
(7) *$\overline{\rho}$ is modular.*

Proof. By Proposition 3.2 of [Cal12], we may find F and $\overline{\rho}$ satisfying all but the last two conditions. By Proposition 8.2.1 of [Sno09], there is a finite Galois extension F'/F in which all places above p split completely such that $\overline{\rho}|_{G_{F'}}$ is modular. If we make a further quadratic extension (linearly disjoint from $\overline{F}^{\ker \overline{\rho}}$ over F, and in which the primes above p split completely), if necessary, to ensure that $[F':\mathbb{Q}]$ is even, and replace F with F', then the result follows. $\square$

For the remainder of the subsection, we follow the arguments of [Kis09a, §2.2] very closely, and we do our best to conform to the notation used there.

We choose a finite set S of finite places of F, containing all the places $v|p$ and at least one other place. Using [Kis09a, Lem. 2.2.1], we can and do choose S so that conditions (1)–(4) of [Kis09a, §2.2] hold.

We denote by D the quaternion algebra with centre F which is ramified at all infinite places of F and unramified at all finite places (so the set Σ considered in [Kis09a, §2.1.1] is empty). We fix a maximal order $\mathcal{O}_D$ of D, for each finite place v we choose an isomorphism $(\mathcal{O}_D)_v \xrightarrow{\sim} M_2(\mathcal{O}_{F_v})$, and we define $U = \prod_v U_v \subset (D \otimes_F \mathbb{A}_F^\infty)^\times$ to be the following compact open subgroup of $\prod_v (\mathcal{O}_D)_v^\times$: if $v|p$ or $v \notin S$ then $U_v = (\mathcal{O}_D)_v^\times$, while if $v \in S$ but $v \nmid p$, then U_v consists of the matrices which are upper-triangular and unipotent modulo ϖ_v, where ϖ_v is a uniformiser of F_v. The subgroup U is sufficiently small in the sense of [Kis09a, §2.1.1].

We write Σ_p for the set of places $v|p$ of F (since Σ is empty, this is consistent with [Kis09a]). For each $v|p$, we let $R_v^{\square,\psi}(\overline{\rho}|_{G_{F_v}})$ denote the universal framed deformation $\mathcal{O}$-algebra for $\overline{\rho}|_{G_{F_v}}$ with determinant $\psi\varepsilon$, and we define

$$R_\infty := \widehat{\otimes}_{v|p,\mathcal{O}} R_v^{\square,\psi}(\overline{\rho}|_{G_{F_v}})[[x_1, \dots, x_g]],$$

where $x_1, \dots, x_g$ are formal variables, and the integer g is chosen as in [Kis09a, Prop. 2.2.4].

For each place $v|p$ of F, we choose integers a_v, b_v with $b_v \geq 0$, together with an inertial type τ_v, and let $*$ be either cr or nothing (the same choice of $*$ being made for all $v|p$). We assume that $\det \tau_v = \varepsilon^{1-2a_v-b_v}\psi|_{I_{\mathbb{Q}_p}}$ for each v. We write[2]

$$\bar{R}_\infty := \widehat{\otimes}_{v|p,\mathcal{O}} R_*^{\square,\psi}(a_v, b_v, \tau_v, \overline{\rho}|_{G_{F_v}})[[x_1, \dots, x_g]].$$

If we write $W_\sigma := \otimes_{v|p} L^*_{a_v,b_v,\tau_v}$, then W_σ is a finite free $\mathcal{O}$-module with an action of $\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})$, and the quotient $W_\sigma/\pi W_\sigma$ is then a representation of $\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})$ over $\mathbb{F}$. Fix a Jordan–Hölder filtration

$$0 = L_0 \subset \cdots \subset L_s = W_\sigma/\pi W_\sigma$$

of $W_\sigma/\pi W_\sigma$ by $\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})$-subrepresentations. If we write $\sigma_i := L_i/L_{i-1}$, then $\sigma_i \cong \otimes_{v|p} \det^{m_{v,i}} \otimes \mathrm{Sym}^{n_{v,i}} \mathbb{F}^2$ for some uniquely determined integers $m_{v,i} \in \{0, \dots, p-2\}$ and $n_{v,i} \in \{0, \dots, p-1\}$.

The patching construction of [Kis09a, § 2.2.5] then gives, for some integer denoted by $h+j$ in [Kis09a], and formal variables $y_1, \dots, y_{h+j}$,

- an $(\bar{R}_\infty, \mathcal{O}[[y_1, \dots, y_{h+j}]])$-bimodule M_∞, finite free over $\mathcal{O}[[y_1, \dots, y_{h+j}]]$, and
- a filtration of $M_\infty/\pi M_\infty$ by $(\bar{R}_\infty/\pi, \mathbb{F}[[y_1, \dots, y_{h+j}]])$-bimodules, say

$$0 = M_\infty^0 \subset M_\infty^1 \subset \cdots \subset M_\infty^s = M_\infty/\pi M_\infty,$$

  such that each $M_\infty^i/M_\infty^{i-1}$ is a finite free $\mathbb{F}[[y_1, \dots, y_{h+j}]]$-module, and such that
- the isomorphism class of $M_\infty^i/M_\infty^{i-1}$ as an $(R_\infty, \mathbb{F}[[y_1, \dots, y_{h+j}]])$-bimodule depends only on the isomorphism class of σ_i as a $\prod_{v|p} \mathrm{GL}_2(\mathcal{O}_{F_v})$-module, and not on the choices of a_v, b_v and τ_v. (It is immediate from the finiteness argument used in patching that this can be achieved for any finite collection

[2] The ring $\bar{R}_\infty$ depends upon the particular choices of a_v, b_v, τ_v and $*$, although (following [Kis09a]) we do not indicate this in the notation.

of tuples (a_v, b_v, τ_v), and since there are only countably many tuples, a diagonalization argument allows us to assume independence for all tuples.)

For any Serre weight $\sigma_{m,n}$, we now write ψ_{cr} for a crystalline character lifting $\omega^{-1}\det\bar{r}$, and we define (in the notation of Remark 3.1.2)

$$\mu_{m,n}(\bar{r}) := e(R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m}, n, \mathbb{1}, \bar{r})/\pi)$$

and

$$\mathcal{C}_{m,n} := Z(R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m}, n, \mathbb{1}, \bar{r})/\pi).$$

Proof of Theorem 3.1.6. Fix a, b, τ and $*$, and set $a_v = a$, $b_v = b$, $\tau_v = \tau$ for each $v|p$. Lemma 2.2.11 of [Kis09a] (together with its proof) shows that $e(\bar{R}_\infty/\pi) \geq e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi)$, and the following conditions are equivalent.

(1) M_∞ is a faithful $\bar{R}_\infty$-module.
(2) M_∞ is a faithful $\bar{R}_\infty$-module which has rank 1 at all generic points of $\bar{R}_\infty$.[3]
(3) $e(\bar{R}_\infty/\pi) = e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi)$.
(4) $e(\bar{R}_\infty/\pi) \leq e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi)$.

(Note that the proof of [Kis09a, Lem. 2.2.11], as written, literally applies only when $a = 0$, but in fact it goes through unchanged for any value of a.) Furthermore, the argument of Corollary 2.2.17 of [Kis09a] (which uses the p-adic Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$, together with the weight part of Serre's conjecture) establishes the inequality $e\big(\bar{R}_\infty/\pi) \leq e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi\big)$, so that in fact all of the conditions (1)–(4) actually hold. (This is where our hypothesis regarding $\bar{r}$ is used.)

Let $\bar{R}^i_\infty := \big(\widehat{\otimes}_{v|p} R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m}_{v,i}, n_{v,i}, \mathbb{1}, \bar{\rho}|_{G_{F_v}})/\pi\big)[[x_1, \ldots, x_g]]$, regarded as a quotient of R_∞. Since the isomorphism class of $M^i_\infty/M^{i-1}_\infty$ as an R_∞-module depends only on the isomorphism class of σ_i, one sees as in the proof of Lemma 2.2.13 of [Kis09a] that the action of R_∞ on $M^i_\infty/M^{i-1}_\infty$ factors through $\bar{R}^i_\infty$. (In brief: it is enough to check for each place $v_0|p$ that the action of $R^{\square,\psi}_{v_0}/\pi$ factors through $R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m}_{v_0,i}, n_{v_0,i}, \mathbb{1}, \bar{\rho}|_{G_{F_{v_0}}})/\pi$. In order to do this, one applies the construction with $a_v = \widetilde{m}_{v_0,i}$, $b_v = n_{v_0,i}$, τ_{v_0} an appropriate scalar type, and the other τ_v chosen so that $\sigma_{m_{v,i},n_{v,i}}$ is a Jordan–Hölder factor of $L_{a_v,b_v,\tau_v} \otimes_{\mathcal{O}} \mathbb{F}$ for $v \neq v_0$.) Furthermore, the R_∞-module $M^i_\infty/M^{i-1}_\infty$ is supported on all of $\mathrm{Spec}\,\bar{R}^i_\infty$, by the results of Section 4.6 of [Gee11] (cf. the final paragraph of the proof of [Kis09a, Prop. 2.2.15]). Finally, we note that $\bar{R}^i_\infty$ is generically reduced (see the proof of [Kis09a, Prop. 2.2.15]).

Lemma 2.2.14 shows that

$$Z(\bar{R}_\infty/\pi) = \Big(\prod_{v|p} Z(R^{\square,\psi}_{*}(a, b, \tau, \bar{r})/\pi)\Big) \times Z(\mathrm{Spec}\,\mathbb{F}[[x_1, \ldots, x_g]]), \tag{3.2.2}$$

while, identifying each of the modules $M_\infty/\pi M_\infty$, $M^i_\infty/M^{i-1}_\infty$, etc., with the corresponding sheaves on $\mathrm{Spec}\,R_\infty/\pi$ that they give rise to, we compute that

$$\begin{aligned} Z(M_\infty/\pi M_\infty) &= \sum_i Z(M^i_\infty/M^{i-1}_\infty) \\ &\geq \sum_i Z(\mathrm{Spec}\,\bar{R}^i_\infty/\pi) = \Big(\prod_{v|p}\sum_{m,n} a^*_{m,n}\mathcal{C}_{m,n}\Big) \times Z(\mathrm{Spec}\,\mathbb{F}[[x_1, \ldots, x_g]]), \end{aligned} \tag{3.2.3}$$

[3] We remark that there is a typo in this part of the statement of [Kis09a, Lem. 2.2.11]; R_∞ is written there, rather than $\bar{R}_\infty$.

the first equality following by Lemma 2.2.7, the inequality following from the fact, noted above, that the support of $M_\infty^i/M_\infty^{i-1}$ coincides with the generically reduced closed subscheme $\operatorname{Spec}\bar{R}_\infty^i$ of $\operatorname{Spec}R_\infty/\pi$, and the second equality following from another application of Lemma 2.2.14. Also, since M_∞ is π-torsion free, and generically free of rank one over each component of $\operatorname{Spec}\bar{R}_\infty$ (condition (2) above), Proposition 2.2.13 shows that

$$Z(\bar{R}_\infty/\pi) = Z(M_\infty/\pi M_\infty). \tag{3.2.4}$$

Putting these computations together, we find that

$$\Big(\prod_{v|p} Z(R_*^{\Box,\psi}(a,b,\tau,\bar{r})/\pi)\Big) \times Z(\operatorname{Spec}\mathbb{F}[[x_1,\dots,x_g]]) \overset{(3.2.2)}{=} Z(\bar{R}_\infty/\pi)$$
$$\overset{(3.2.4)}{=} Z(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi) \overset{(3.2.3)}{\geq} \Big(\prod_{v|p}\sum_{m,n} a^*_{m,n}\mathcal{C}_{m,n}\Big) \times Z(\operatorname{Spec}\mathbb{F}[[x_1,\dots,x_g]]).$$

However, if we apply Lemma 2.2.8 (with y the closed point of $\operatorname{Spec}\bar{R}_\infty/\pi$) to pass to the corresponding multiplicities, then this inequality on cycles gives a corresponding inequality on multiplicities, which is in fact an equality, by Lemma 2.3.1 of [Kis09a]. Thus this inequality of cycles is an equality (note that any non-zero cycle must have non-zero multiplicity at the unique closed point of $\operatorname{Spec}\bar{R}_\infty/\pi$), and we deduce that

$$\prod_{v|p} Z(R_*^{\Box,\psi}(a,b,\tau,\bar{r})/\pi) = \prod_{v|p}\sum_{m,n} a^*_{m,n}\mathcal{C}_{m,n},$$

and thus that

$$Z(R_*^{\Box,\psi}(a,b,\tau,\bar{r})/\pi) = \sum_{m,n} a^*_{m,n}\mathcal{C}_{m,n},$$

as required. $\square$

3.2.5. **Remark.** Even if $\bar{r} \sim \begin{pmatrix}\omega\chi & * \\ 0 & \chi\end{pmatrix}$ for some χ, it is presumably possible to use the above arguments to show that our geometric Breuil–Mézard conjecture is equivalent to the usual one, and that both are equivalent to the equivalent conditions of Lemma 2.2.11 of [Kis09a]. We have not done so, because parity issues with Hilbert modular forms make the argument rather longer than one would wish, and in any case we prove a similar statement in far greater generality in Section 5 (see Theorem 5.5.2, which together with Lemma 4.3.1 shows the equivalence of our geometric conjecture with the usual one). (Note that since $p > 2$ all the Serre weights occurring in the reduction mod p of $L_{a,b,\tau}$ have the same parity, so it is possible to circumvent parity problems by twisting, but the details are a little unpleasant to write out.)

3.3. **Analysis of components.** For a given $\bar{r}$, there may be several different Serre weights $\sigma_{m,n}$ for which $\mu_{m,n}(\bar{r}) \neq 0$. We now examine the different cycles $\mathcal{C}_{m,n}$.

We write $W(\bar{r})$ for the set of $\sigma_{m,n}$ for which $\mu_{m,n}(\bar{r}) \neq 0$ (that is: the set of Serre weights for $\bar{r}$). The only cases where $W(\bar{r})$ contains more than one element are as follows (throughout it is understood that we always impose the conditions that $0 \leq m \leq p-2$ and $0 \leq n \leq p-1$):

- If $\bar{r}$ is irreducible, say
$$\bar{r}|_{I_{\mathbb{Q}_p}} \sim \omega^m \otimes \begin{pmatrix} \omega_2^{n+1} & 0 \\ 0 & \omega_2^{p(n+1)} \end{pmatrix},$$
then $W(\bar{r}) = \{\sigma_{m,n}, \sigma_{m+n,p-1-n}\}$.
- If $\bar{r}$ is reducible but indecomposable, with
$$\bar{r}|_{I_{\mathbb{Q}_p}} \sim \begin{pmatrix} \omega^{m+1} & * \\ 0 & \omega^m \end{pmatrix}$$
a peu ramifiée extension, then $W(\bar{r}) = \{\sigma_{m,0}, \sigma_{m,p-1}\}$.
- If $\bar{r}$ is reducible and decomposable, with
$$\bar{r}|_{I_{\mathbb{Q}_p}} \sim \begin{pmatrix} \omega^{m+n+1} & 0 \\ 0 & \omega^m \end{pmatrix}$$
such that $n \le p-3$, then
  - if $0 < n < p-3$, we have $W(\bar{r}) = \{\sigma_{m,n}, \sigma_{m+n+1,p-3-n}\}$;
  - if $n = p-3$ and $p > 3$, we have $W(\bar{r}) = \{\sigma_{m,p-3}, \sigma_{m-1,0}, \sigma_{m-1,p-1}\}$;
  - if $p = 3$ and $n = 0$, we have $W(\bar{r}) = \{\sigma_{m,0}, \sigma_{m,2}, \sigma_{m+1,0}, \sigma_{m+1,2}\}$.

The relationships between the cycles $\mathcal{C}_{m,n}$ are as follows. Note in particular that by Theorem 3.1.6, if $\bar{r} \not\sim \begin{pmatrix} \omega\chi & * \\ 0 & \chi \end{pmatrix}$ for any χ, then the assumption in the following Proposition is automatic.

3.3.1. **Proposition.** *Assume that Conjecture* 3.1.4 *holds for* $\bar{r}$. *Then the cycles* $\mathcal{C}_{m,n}$ *have disjoint support, except for the cycles* $\mathcal{C}_{m,0}$ *and* $\mathcal{C}_{m,p-1}$ *when both are non-zero. In this latter case there is an equality* $\mathcal{C}_{m,0} = \mathcal{C}_{m,p-1}$, *except if* $\bar{r}$ *is a twist of a (possibly split) peu ramifiée extension of the trivial character by the mod* p *cyclotomic character, in which case* $\mathcal{C}_{m,p-1}$ *is the sum of two irreducible cycles, one of which is* $\mathcal{C}_{m,0}$.

Proof. It is presumably possible to establish this in most cases via direct computations with Fontaine-Laffaille theory; however, we take the opportunity to use our geometric formulation of the Breuil–Mezard conjecture. We begin by noting that it follows from Corollary 1.7.14 of [Kis09a] that the cycles $\mathcal{C}_{m,n}$ are irreducible, except in the cases that $n = p-1$ and $\bar{r}|_{I_{\mathbb{Q}_p}} \sim \begin{pmatrix} \omega^{m+1} & * \\ 0 & \omega^m \end{pmatrix}$ is peu ramifiée, or $n = p-2$ and $\bar{r} \sim \omega^m \otimes \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}$ for distinct unramified characters μ_1 and μ_2. Thus to prove the claimed disjointness of cycles, it is enough to prove that the cycles are not equal.

We first consider the cases where $\mathcal{C}_{m,0}$ and $\mathcal{C}_{m,p-1}$ are nonzero. For notational simplicity, we make a twist by ω^{-m}, and thus further assume that $m = 0$. Consider the trivial inertial type $\mathbb{1}$. Examining Henniart's appendix to [BM02], we see that $\sigma(\mathbb{1}) = \mathrm{St}$, the Steinberg representation, while $\sigma^{\mathrm{cr}}(\mathbb{1}) = \mathbb{1}$, the trivial representation. The reduction mod p of St is just σ_{p-1}, so we conclude that
$$\mathcal{C}_{0,p-1} = Z(R^{\square,\psi_{\mathrm{cr}}}(\widetilde{0}, 0, \mathbb{1}, \bar{r})/\pi),$$
$$\mathcal{C}_{0,0} = Z(R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{0}, 0, \mathbb{1}, \bar{r})/\pi).$$

By definition, $R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{0},0,\mathbb{1},\bar{r})$ is a quotient of $R^{\square,\psi}(0,0,\mathbb{1},\bar{r})$, and as they have the same dimension, we see that $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{0},0,\mathbb{1},\bar{r})$ is a union of irreducible components of $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}(\widetilde{0},0,\mathbb{1},\bar{r})$.

It is easy to see that the two rings are actually equal unless $\bar{r}$ is a twist of a peu ramifié extension of 1 by ω, because the only non-crystalline semistable crystalline representations with Hodge–Tate weights 0 and 1 are unramified twists of extensions of 1 by ε, and so it suffices to check, in the case when $\bar{r}$ *is* a twist of a peu ramifié extension of 1 by ω, that $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}(\widetilde{0},0,\mathbb{1},\bar{r})/\pi)$ has two distinct components. In the case that the extension is non-split, the framed deformation ring is formally smooth over the deformation ring, and the relevant rings are computed in Theorem 5.3.1(i) of [BM02]; in particular, they do verify that the Specs of their reductions mod π contain two distinct components.

We now explain another way to see that $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}(\widetilde{0},0,\mathbb{1},\bar{r})/\pi)$ consists of two distinct components, which works equally well in the case when $\bar{r}$ is split. Namely, a two-dimensional crystalline representation with Hodge–Tate weights 0 and 1 with reducible reduction is necessarily an extension of characters, whose restrictions to I_p are trivial and cyclotomic respectively, and it is uniquely determined by its associated pseudo-representation. The cycle $\mathcal{C}_{0,0}$ is thus directly seen to be irreducible, and it has non-trivial image in the associated pseudo-deformation space. On the other hand, as we already observed, a genuinely semi-stable two-dimensional deformation of $\bar{r}$ with Hodge–Tate weights 0 and 1 is necessarily a twist of an extension of the trivial character by the cyclotomic character, with the possible twist being uniquely determined (since we have fixed the determinant to be $\psi\varepsilon$). Such extensions are determined by their $\mathcal{L}$-invariant, and so one can give an explicit description of the Zariski closure of the space of genuinely semi-stable deformations, and show that this closure, as well as its reduction mod π, is irreducible. Moreover, its reduction mod π does not coincide with $\mathcal{C}_{0,0}$, since its image in the associated pseudo-deformation space is simply the closed point. Thus we have shown that $\mathcal{C}_{0,p-1}$ is the sum of two distinct irreducible components.

Consider next the case that $\bar{r}|_{I_{\mathbb{Q}_p}} \sim \omega^m \otimes \begin{pmatrix} \omega_2^{n+1} & 0 \\ 0 & \omega_2^{p(n+1)} \end{pmatrix}$ with $0 < n < p-1$. We need to show that $\mathcal{C}_{m,n} \neq \mathcal{C}_{m+n,p-1-n}$. Consider the inertial type $\widetilde{\omega}^{m+n} \oplus \widetilde{\omega}^m$, and choose ψ so that $\psi|_{I_{\mathbb{Q}_p}} = \varepsilon\widetilde{\omega}^{2m+n}$. Then by for example Lemmas 3.1.1 and 4.2.4 of [CDT99], the semisimplification of the reduction modulo p of $\sigma(\widetilde{\omega}^{m+n} \oplus \widetilde{\omega}^m)$ (the representation of $\mathrm{GL}_2(\mathbb{Z}_p)$ associated to $\widetilde{\omega}^{m+n} \oplus \widetilde{\omega}^m$ by Henniart's inertial local Langlands correspondence, which is the inflation to $\mathrm{GL}_2(\mathbb{Z}_p)$ of a principal series representation of $\mathrm{GL}_2(\mathbb{F}_p)$) has Jordan–Hölder factors $\sigma_{m,n}$ and $\sigma_{m+n,p-1-n}$, so that we have

$$Z(R^{\square,\psi}(0,1,\widetilde{\omega}^{m+n} \oplus \widetilde{\omega}^m,\bar{r})/\pi) = \mathcal{C}_{m,n} + \mathcal{C}_{m+n,p-1-n}.$$

By Theorem 6.22 of [Sav05] (noting that the framed deformation ring is formally smooth over the deformation ring, since $\bar{r}$ is irreducible), we have $R^{\square,\psi}(0,1,\widetilde{\omega}^{m+n} \oplus \widetilde{\omega}^m,\bar{r})/\pi \cong \mathbb{F}[[U,V,W,X,Y]]/(XY)$, so $\mathcal{C}_{m,n} \neq \mathcal{C}_{m+n,p-1-n}$, as required.

It remains to treat the cases where $\bar{r} \sim \overline{\chi}_1 \oplus \overline{\chi}_2$ is a direct sum of two distinct characters. In this case, one knows that all of the relevant lifting rings $R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m},n,\mathbb{1},\bar{r})$ are in fact ordinary lifting rings (cf. Corollary 1.7.14 and Remark 1.7.16 of [Kis09a]). It thus suffices to note that the cycles that we have to prove are distinct correspond respectively to liftings which contain a submodule lifting $\overline{\chi}_1$ or which contain a submodule lifting $\overline{\chi}_2$, and since $\overline{\chi}_1 \neq \overline{\chi}_2$, it is easy to see that the cycles are distinct. $\square$

Proof of Proposition 1.1.3. The assertions about the number of components of $\mathcal{C}_{m,n}$ follow from Proposition 3.3.1 above together with Corollary 1.7.14 of [Kis09a], and the disjointness of the supports of the $\mathcal{C}_{m,n}$ follows from Proposition 3.3.1. $\square$

3.4. **Comparison with the results of** [BM12]**.** In [BM12], Breuil and Mézard construct, for generic $\bar{r}$, a correspondence between the irreducible components of $\operatorname{Spec} R^{\square,\psi}(a,b,\tau,\bar{r})/\pi$ and the Serre weights for $\bar{r}$ which appear with positive multiplicity in the mod p reduction of $(\det^a \otimes \mathrm{Sym}^b E^2) \otimes_E \sigma(\tau)$. The definition of the correspondence is given in [BM12, Thm. 1.5]: namely, if $\mathfrak{a}$ is a generic point of $\operatorname{Spec} R^{\square,\psi}(a,b,\tau,\bar{r})/\pi$, then the associated Serre weight is the $\mathrm{GL}_2(\mathbb{Z}_p)$-socle of a certain $\mathrm{GL}_2(\mathbb{Q}_p)$-representation obtained from the universal Galois representation into $\mathrm{GL}_2\big(R^{\square,\psi}(a,b,\tau,\bar{r})/(\pi,\mathfrak{a})\big)$ via the p-adic local Langlands correspondence.

According to our geometric formulation of the Breuil–Mézard conjecture, the component $\overline{\{\mathfrak{a}\}}$ of $\operatorname{Spec} R^{\square,\psi}(a,b,\tau,\bar{r})/\pi$ is equal to $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m},n,\mathbb{1},\bar{r})/\pi)$, for some Serre weight $\sigma_{m,n}$ of $\bar{r}$. Since the p-adic local Langlands correspondence is functorial, we find that the $\mathrm{GL}_2(\mathbb{Z}_p)$-socle of the $\mathrm{GL}_2(\mathbb{Q}_p)$-representation attached to $\mathfrak{a}$, thought of as *a* generic point of $\operatorname{Spec} R^{\square,\psi}(a,b,\tau,\bar{r})/\pi$, is the same as the $\mathrm{GL}_2(\mathbb{Z}_p)$-socle of the $\mathrm{GL}_2(\mathbb{Q}_p)$-representation attached to $\mathfrak{a}$, thought of as *the* generic point of $\operatorname{Spec} R^{\square,\psi_{\mathrm{cr}}}_{\mathrm{cr}}(\widetilde{m},n,\mathbb{1},\bar{r})/\pi)$. But in this latter case, the $\mathrm{GL}_2(\mathbb{Z}_p)$-socle in question is equal to $\sigma_{m,n}$, as follows from the fact that the correspondence of [BM12] is compatible with the original conjecture of [BM02]. This shows that the correspondence of [BM12] is precisely the correspondence $\mathcal{C}_{m,n} \mapsto \sigma_{m,n}$.

4. The Breuil–Mézard Conjecture for GL_n

4.1. **The numerical conjecture.** We begin by formulating a natural generalisation of the Breuil–Mézard conjecture for n-dimensional representations. We now fix the notation we will use for the rest of the paper, which differs in some respects from that of Section 3, but is closer to that used in the literature on automorphy lifting theorems for unitary groups. We remind the reader that for the rest of the paper, our convention on Hodge–Tate weights is that the Hodge–Tate weight of ε is -1. Let $K/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathcal{O}_K$ and residue field k, let $E/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$, uniformiser π and residue field $\mathbb{F}$, and let $\bar{r} : G_K \to \mathrm{GL}_n(\mathbb{F})$ be a continuous representation. Assume that E is sufficiently large, and in particular that E contains the images of all embeddings $K \hookrightarrow \overline{\mathbb{Q}}_p$. Let $\mathbb{Z}^n_+$ denote the set of tuples $(\lambda_1,\ldots,\lambda_n)$ of integers with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. For any $\lambda \in \mathbb{Z}^n_+$, view λ as a dominant character of the algebraic group $\mathrm{GL}_{n/\mathcal{O}}$ in the usual way, and let M'_λ be the algebraic $\mathcal{O}_K$-representation of GL_n given by

$$M'_\lambda := \mathrm{Ind}_{B_n}^{\mathrm{GL}_n}(w_0\lambda)_{/\mathcal{O}_K}$$

where B_n is the Borel subgroup of upper-triangular matrices of GL_n, and w_0 is the longest element of the Weyl group (see [Jan03] for more details of these notions, and note that M'_λ has highest weight λ). Write M_λ for the $\mathcal{O}_K$-representation of $\mathrm{GL}_n(\mathcal{O}_K)$ obtained by evaluating M'_λ on $\mathcal{O}_K$. For any $\lambda \in (\mathbb{Z}^n_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$ we write L_λ for the $\mathcal{O}$-representation of $\mathrm{GL}_n(\mathcal{O}_K)$ defined by

$$\otimes_{\tau:K\hookrightarrow E} M_{\lambda_\tau} \otimes_{\mathcal{O}_K,\tau} \mathcal{O}.$$

Given any $a \in \mathbb{Z}^n_+$ with $p-1 \geq a_i - a_{i+1}$ for all $1 \leq i \leq n-1$, we define the k-representation P_a of $\mathrm{GL}_n(k)$ to be the representation obtained by evaluating $\mathrm{Ind}_{B_n}^{\mathrm{GL}_n}(w_0 a)_k$ on k, and let N_a be the irreducible sub-k-representation of P_a generated by the highest weight vector (that this is indeed irreducible follows for example from II.2.8(1) of [Jan03] and the appendix to [Her09]). We say that an element $a \in (\mathbb{Z}^n_+)^{\mathrm{Hom}(k,\mathbb{F})}$ is a *Serre weight* if

- for each $\sigma \in \mathrm{Hom}(k, \mathbb{F})$ and each $1 \leq i \leq n-1$ we have
$$p - 1 \geq a_{\sigma,i} - a_{\sigma,i+1},$$
- and for each σ we have $0 \leq a_{\sigma,n} \leq p-1$, and not all $a_{\sigma,n} = p-1$.

If $a \in (\mathbb{Z}^n_+)^{\mathrm{Hom}(k,\mathbb{F})}$ is a Serre weight then we define an irreducible $\mathbb{F}$-representation F_a of $\mathrm{GL}_n(k)$ by
$$F_a := \otimes_{\tau \in \mathrm{Hom}(k,\mathbb{F})} N_{a_\tau} \otimes_{k,\tau} \mathbb{F}.$$
The representations F_a are absolutely irreducible and pairwise non-isomorphic, and every irreducible $\mathbb{F}$-representation of $\mathrm{GL}_n(k)$ is of the form F_a for some a (see for example the appendix to [Her09]). We say that an element $\lambda \in (\mathbb{Z}^n_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$ is a *lift* of a Serre weight $a \in (\mathbb{Z}^n_+)^{\mathrm{Hom}(k,\mathbb{F})}$ if for each $\sigma \in \mathrm{Hom}(k, \mathbb{F})$ there is an element $\tau \in \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$ lifting σ such that $\lambda_\tau = a_\sigma$, and for all other $\tau' \in \mathrm{Hom}_{\mathbb{Q}_p}(K, E)$ lifting σ we have $\lambda_{\tau'} = 0$. We have a partial ordering $\leq$ on Serre weights, where $b \leq a$ if and only if $a - b$ is a sum of (positive) simple roots.

4.1.1. **Lemma.** *If λ is a lift of a then $L_\lambda \otimes_{\mathcal{O}} \mathbb{F}$ has socle F_a, and every other Jordan–Hölder factor of $L_\lambda \otimes_{\mathcal{O}} \mathbb{F}$ is of the form F_b with $b < a$.*

Proof. This follows from sections 5.8 and 5.9 of [Hum06] (noting that the orderings $\leq$ and $\leq_{\mathbb{Q}}$ coincide for GL_n). □

Let $\tau : I_K \to \mathrm{GL}_n(E)$ be a representation with open kernel which extends to W_K, and take $\lambda \in (\mathbb{Z}^n_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$. Let $\bar{r} : G_K \to \mathrm{GL}_n(\mathbb{F})$ be a continuous representation. If E'/E is a finite extension, we say that a potentially crystalline representation $\rho : G_K \to \mathrm{GL}_n(E')$ has *Hodge type* λ if for each $\tau : K \hookrightarrow E$,
$$\mathrm{HT}_\tau(\rho) = \{\lambda_{\tau,1} + n - 1, \lambda_{\tau,2} + n - 2, \ldots, \lambda_{\tau,n}\}.$$
We say that ρ has *inertial type* τ if the restriction to I_K of the Weil–Deligne representation associated to ρ is equivalent to τ.

4.1.2. **Proposition.** *For each λ, τ there is a unique quotient $R^\square_{\bar{r},\lambda,\tau}$ of the universal lifting $\mathcal{O}$-algebra $R^\square_{\bar{r}}$ for $\bar{r}$ with the following properties.*

(1) *$R^\square_{\bar{r},\lambda,\tau}$ is reduced and p-torsion free, and $R^\square_{\bar{r},\lambda,\tau}[1/p]$ is formally smooth and equidimensional of dimension $n^2 + [K : \mathbb{Q}_p]n(n-1)/2$.*
(2) *If E'/E is a finite extension, then an $\mathcal{O}$-algebra homomorphism $R^\square_{\bar{r}} \to E'$ factors through $R^\square_{\bar{r},\lambda,\tau}$ if and only if the corresponding representation $G_K \to \mathrm{GL}_n(E')$ is potentially crystalline of Hodge type λ and inertial type τ.*
(3) *$R^\square_{\bar{r},\lambda,\tau}/\pi$ is equidimensional.*

Proof. All but the final point are proved in [Kis08], and the final statement is a straightforward consequence of the first (cf. the proof of Lemme 2.1 of [BM12]). □

If τ is trivial we will write $R^\square_{\bar{r},\lambda}$ for $R^\square_{\bar{r},\lambda,\tau}$. The Breuil–Mézard conjecture predicts the Hilbert–Samuel multiplicity $e(R^\square_{\bar{r},\lambda,\tau}/\pi)$. In order to state the conjecture, it is first necessary to make a conjecture on the existence of an inertial local Langlands correspondence for GL_n. The following is a folklore conjecture.

4.1.3. **Conjecture.** *If τ is an inertial type, then there is a finite-dimensional smooth irreducible $\overline{\mathbb{Q}}_p$-representation $\sigma(\tau)$ of $\mathrm{GL}_n(\mathcal{O}_K)$ such that if $\widetilde{\tau}$ is any Frobenius-semisimple Weil–Deligne representation of W_K over $\overline{\mathbb{Q}}_p$, then the restriction of $(\mathrm{rec}_p^{-1}(\widetilde{\tau})^\vee)$ to $\mathrm{GL}_n(\mathcal{O}_K)$ contains (an isomorphic copy of) $\sigma(\tau)$ as a subrepresentation if and only if $\widetilde{\tau}|_{I_K} \sim \tau$ and $N = 0$ on $\widetilde{\tau}$. If $p > n$ then $\sigma(\tau)$ is unique up to isomorphism.*

4.1.4. **Remark.** A very similar conjecture is formulated in [Con10], which also proves some partial results in the case $n = 3$. We have formulated this conjecture only for representations with $N = 0$, as we will only formulate our generalisations of the Breuil–Mézard conjecture for potentially crystalline representations, in order to avoid complications in the global arguments of Section 5. We expect that a "semistable" version of the conjecture will also be valid, where one removes the conclusion that $N = 0$ on τ, but adds the requirement that $\mathrm{rec}_p^{-1}(\widetilde{\tau}) \otimes |\det|^{(n-1)/2}$ be generic.

Conjecture 4.1.3 is proved in the case $n = 2$ by Henniart in the appendix to [BM02]. It has been proved for any n for supercuspidal representations by Paškūnas ([Paš05]), but to the best of our knowledge it is open in general. However, the important point for us is the existence of $\sigma(\tau)$, rather than its uniqueness, and this is known in general. The following is a special case of Proposition 6.5.3 of [BC09] (see also [Con12]).

4.1.5. **Theorem.** *If τ is an inertial type, then there is a finite-dimensional smooth irreducible $\overline{\mathbb{Q}}_p$-representation $\sigma(\tau)$ of $\mathrm{GL}_n(\mathcal{O}_K)$ such that if $\widetilde{\tau}$ is any pure Frobenius-semisimple Weil–Deligne representation of W_K over $\overline{\mathbb{Q}}_p$, then the restriction of $(\mathrm{rec}_p^{-1}(\widetilde{\tau}) \otimes |\det|^{(n-1)/2})$ to $\mathrm{GL}_n(\mathcal{O}_K)$ contains (an isomorphic copy of) $\sigma(\tau)$ as a subrepresentation if and only if $\widetilde{\tau}|_{I_K} \sim \tau$ and $N = 0$ on $\widetilde{\tau}$.*

Enlarging E if necessary, we may assume that $\sigma(\tau)$ is defined over E. Since it is a finite-dimensional representation of the compact group $\mathrm{GL}_n(\mathcal{O}_K)$, it contains a $\mathrm{GL}_n(\mathcal{O}_K)$-stable $\mathcal{O}$-lattice L_τ. Set $L_{\lambda,\tau} := L_\tau \otimes_{\mathcal{O}} L_\lambda$, a finite free $\mathcal{O}$-module with an action of $\mathrm{GL}_n(\mathcal{O}_K)$. Then we may write

$$(L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_a F_a^{n_a},$$

where the sum runs over the Serre weights $a \in (\mathbb{Z}_+^n)^{\mathrm{Hom}(k,\mathbb{F})}$, and the n_a are non-negative integers. Then the generalised Breuil–Mézard conjecture is the following.

4.1.6. **Conjecture.** *There exist integers $\mu_a(\bar{r})$ depending only on $\bar{r}$ and a such that $e(R^{\square}_{\bar{r},\lambda,\tau}/\pi) = \sum_a n_a \mu_a(\bar{r})$.*

4.1.7. **Remark.**

(1) Assuming the conjecture, the integers $\mu_a(\bar{r})$ may be computed recursively as follows. Let λ_a be a lift of a. If a is in the lowest alcove, then $L_\lambda \otimes_{\mathcal{O}} \mathbb{F} = F_a$, and we have $\mu_a(\bar{r}) = e(R^{\square}_{\bar{r},\lambda_a}/\pi)$. In general, by Lemma 4.1.1 we see that we may compute $\mu_a(\bar{r})$ given $e(R^{\square}_{\bar{r},\lambda_a}/\pi)$ and the values $\mu_b(\bar{r})$, $b < a$.

(2) The reader might have expected the sum on the right hand side to only be over weights a which are predicted Serre weights for $\bar{r}$. However, according

to the philosophy explained in the introduction to [GK12], the predicted Serre weights a for $\bar{r}$ should be precisely the a for which $\mu_a(\bar{r}) \neq 0$.

(3) The reader might wonder why we have formulated our n-dimensional analogues of the Breuil–Mézard conjectures for liftings without fixed determinant, when the conjecture is more usually stated as in Section 3 for lifts with fixed determinant. The reason is that in the global arguments we will make in Section 5 using unitary groups, it is the lifting rings without fixed determinant that arise naturally. However, the conjectures with and without fixed determinant are actually equivalent, at least as long as $p > n$, as follows from Lemma 4.3.1 below.

4.2. **The geometric conjecture.** We may now state our geometric conjecture, which is entirely analogous to Conjecture 3.1.4.

4.2.1. **Conjecture.** *For each Serre weight a there is a cycle $\mathcal{C}_a$ depending only on $\bar{r}$ and a such that*

$$Z(R^{\square}_{\bar{r},\lambda,\tau}/\pi) = \sum_a n_a \mathcal{C}_a.$$

4.2.2. **Remark.** Again, if one assumes the conjecture then one can inductively compute the cycles $\mathcal{C}_a$ in terms of the cycles $Z(R^{\square}_{\bar{r},\lambda_a}/\pi)$.

4.3. **Twisting by characters.** We now establish a lemma which implies in particular that the Breuil–Mézard conjecture for liftings without a fixed determinant (as we have formulated it above) is equivalent to the analogous conjecture for liftings with a fixed determinant (at least if $p > n$). Note that, for a given inertial type τ and Hodge type λ, there is a character $\psi_{\lambda,\tau} : I_K \to \mathcal{O}^\times$ such that any lift ρ of $\bar{r}$ of Hodge type λ and inertial type τ necessarily has $\det \rho|_{I_K} = \psi_{\lambda,\tau}$. Let $\psi : G_K \to \mathcal{O}^\times$ be a character such that $\overline{\psi} = \det \bar{r}$ and $\psi|_{I_K} = \psi_{\lambda,\tau}$. We let $R^{\square,\psi}_{\bar{r},\lambda,\tau}$ denote the quotient of $R^{\square}_{\bar{r},\lambda,\tau}$ corresponding to lifts with determinant ψ. In the case $n = 2$ and $K = \mathbb{Q}_p$ this is [BM02, 2.2.2.9].

4.3.1. **Lemma.** *Suppose that $p > n$. Then $R^{\square}_{\bar{r},\lambda,\tau} \cong R^{\square,\psi}_{\bar{r},\lambda,\tau}[[X]]$.*

Proof. Consider a finite extension E'/E and a point $x : R^{\square}_{\bar{r},\lambda,\tau} \to E'$ with corresponding representation ρ'. Let $\theta = \psi^{-1} \det \rho'$, so that θ is an unramified character with trivial reduction. Since $p > n$, we see that θ has an $\mathcal{O}_{E'}$-valued n-th root. From this observation is it easy to see that if $\rho^\psi : G_K \to \mathrm{GL}_n(R^{\square,\psi}_{\bar{r},\lambda,\tau})$ is the universal lifting of determinant ψ, and μ_x is the unramified character taking a Frobenius element to x, then $\rho^\psi \otimes \mu_{1+X} \circ \det : G_K \to \mathrm{GL}_n(R^{\square,\psi}_{\bar{r},\lambda,\tau}[[X]])$ is the universal lifting of arbitrary determinant, as required. $\square$

5. Global patching arguments

Our goal in this section is to employ the Taylor–Wiles–Kisin patching method so as to generalize, to the extent possible, the results of Section 3 to the n-dimensional context. These arguments are essentially the natural n-dimensional generalisation of the arguments of Section 4 of [GK12]. We closely follow the approaches of [Kis09a] and [Tho12] (which in turn follows [BLGG11] and [CHT08]). In particular, in the actual implementation of the patching method we follow [Tho12] very closely, although, because our ultimate interests are local, we will sometimes make stronger global assumptions in order to simplify the arguments; these stronger assumptions can always be achieved in our applications. Before getting to the patching argument itself, we include a number of preliminary subsections in which we briefly recall the necessary background material on automorphic forms and Galois representations, referring the reader to [Tho12] for more details.

5.1. **Basic set-up.** We put ourselves in the setting of Section 4, so that $K/\mathbb{Q}_p$ is a finite extension, and $\bar{r} : G_K \to \mathrm{GL}_n(\mathbb{F})$ is a continuous representation. We also assume from now on that $p > 2$.

As in Section 3, we begin by globalising $\bar{r}$. Since the standard global context in which to study higher-dimensional Galois representations is that of automorphic forms on unitary groups, we briefly recall the various concepts that are required to discuss Galois representations in that setting.

To begin with, we recall from [CHT08] that $\mathcal{G}_n$ denotes the group scheme over $\mathbb{Z}$ defined to be the semidirect product of $\mathrm{GL}_n \times \mathrm{GL}_1$ by the group $\{1, j\}$, which acts on $\mathrm{GL}_n \times \mathrm{GL}_1$ by

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

We have a homomorphism $\nu : \mathcal{G}_n \to \mathrm{GL}_1$, sending (g, μ) to μ and j to -1. We refer the reader to Section 2.1 of [CHT08] for a thorough discussion of $\mathcal{G}_n$, and of the relationship between $\mathcal{G}_n$-valued representations and essentially conjugate self-dual GL_n-valued representations.

5.1.1. **Terminology.** To ease notation, we adopt the following convention with regard to Galois representations with values in $\mathcal{G}_n(\overline{\mathbb{F}}_p)$: if F is an imaginary CM field with maximal totally real subfield F^+, and $\overline{\rho} : G_{F^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$ is a continuous representation with $\overline{\rho}(G_F) \subset \mathrm{GL}_n(\overline{\mathbb{F}}_p) \times \mathrm{GL}_1(\overline{\mathbb{F}}_p)$, then we write $\overline{\rho}|_{G_F}$ for the restriction of $\overline{\rho}$ to G_F, regarded as a representation $G_F \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$, and similarly for $\overline{\rho}(G_{F(\zeta_p)})$ and $\overline{\rho}|_{G_{F_{\widetilde{v}}}}$ (for places $\widetilde{v}$ of F).

Recall that the notion of an *adequate* subgroup of $\mathrm{GL}_n(\overline{\mathbb{F}}_p)$ is defined in [Tho12]. We will not need the details of the definition; some examples of representations whose image is adequate will be constructed in Appendix A.

We now state our basic hypothesis related to the globalization of $\bar{r}$. Namely, we assume that there is an imaginary CM field F with maximal totally real subfield F^+, together with a continuous representation $\overline{\rho} : G_{F^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$, such that

- F/F^+ is unramified at all finite places,
- $[F^+ : \mathbb{Q}]$ is divisible by 4,
- every place $v|p$ of F^+ splits in F,
- $\overline{\rho}$ is automorphic in the sense of Definition 5.3.1 below; in particular, $\nu \circ \overline{\rho} = \overline{\varepsilon}^{1-n}\delta^n_{F/F^+}$, where δ_{F/F^+} is the quadratic character corresponding to F/F^+,
- $\overline{\rho}^{-1}(\mathrm{GL}_n(\overline{\mathbb{F}}_p) \times \mathrm{GL}_1(\overline{\mathbb{F}}_p)) = G_F$,
- $\overline{\rho}$ is unramified at primes $v \nmid p$,
- $\overline{\rho}(G_{F(\zeta_p)})$ is adequate, so that in particular $\overline{\rho}|_{G_F}$ is irreducible,
- $\overline{F}^{\ker \mathrm{ad}\, \overline{\rho}|_{G_F}}$ does not contain $F(\zeta_p)$, and
- for each place $v|p$ of F^+, there is a place $\widetilde{v}$ of F lying over v such that $F_{\widetilde{v}} \cong K$ and $\overline{\rho}|_{G_{F_{\widetilde{v}}}}$ is isomorphic to $\bar{r}$.

We say that such an (F and) $\overline{\rho}$ is a *suitable globalization* of $\bar{r}$.

5.1.2. **Remark.** It will not always be the case that such a representation $\overline{\rho}$ exists, because the assumption that $\overline{\rho}(G_{F(\zeta_p)})$ is adequate implies that $p \nmid n$. On the other hand, if we assume that $p \nmid n$ and that Conjecture A.3 holds for $\bar{r}$, then by Corollary A.7 there is a suitable globalization of $\bar{r}$.

We briefly explain the motivation for the various conditions that we require. The first three ensure the existence of a convenient unitary group on which to work, with the property that it is isomorphic to GL_n at places dividing p. The final condition ensures that $\overline{\rho}$ can be used to study $\bar{r}$. The remaining conditions are imposed in order to use patching constructions of [Tho12]; some of them are imposed in order to simplify these constructions. (When considering the last three conditions, the reader should recall our terminological convention of (5.1.1).)

5.2. **Unitary groups and algebraic automorphic forms.** There is a reductive algebraic group G/F^+ with the following properties (cf. Section 6 of [Tho12]):

- G is an outer form of GL_n, with $G_{/F} \cong \mathrm{GL}_{n/F}$.
- If v is a finite place of F^+, G is quasi-split at v.
- If v is an infinite place of F^+, then $G(F_v^+) \cong U_n(\mathbb{R})$.

As in section 3.3 of [CHT08] we may define a model for G over $\mathcal{O}_{F^+}$. If v is a place of F^+ which splits as ww^c over F, then we have an isomorphism

$$\iota_w : G(\mathcal{O}_{F_v^+}) \xrightarrow{\sim} \mathrm{GL}_n(\mathcal{O}_{F_w}).$$

Let $E/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$ and residue field $\mathbb{F}$, which we assume is chosen to be sufficiently large that $\overline{\rho}$ is valued in $\mathcal{G}_n(\mathbb{F})$. From now on we will often regard $\overline{\rho}|_{G_F}$ as being valued in $\mathrm{GL}_n(\mathbb{F})$, and we write $V_{\mathbb{F}}$ for the underlying $\mathbb{F}$-vector space of $\overline{\rho}|_{G_F}$.

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 $F_p^+ := F \otimes_{\mathbb{Q}} \mathbb{Q}_p$, $\mathcal{O}_{F_p^+} := \mathcal{O}_F \otimes_{\mathbb{Z}} \mathbb{Z}_p$. Let W be a finite $\mathcal{O}$-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 $u_p \in G(\mathcal{O}_{F_p^+})$. We say that U is *sufficiently small* if for all $t \in G(\mathbb{A}_{F^+}^{\infty})$, $t^{-1}G(F^+)t \cap U$ does not contain an element of order p. Let $S(U, W)$ denote the space of algebraic modular forms on G of level U and weight W, i.e. 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 $u \in U$. If U is sufficiently small, then the functor $W \mapsto S(U, W)$ is exact.

For each $v \in S_p$ choose an inertial type $\tau_v : I_{F_v} \to \mathrm{GL}_n(E)$ and a weight $\lambda_v \in (\mathbb{Z}_+^n)^{\mathrm{Hom}_{\mathbb{Q}_p}(F_{\widetilde{v}}, E)}$, and let L_{λ_v,τ_v} be the $\mathcal{O}$-representation of $\mathrm{GL}_n(\mathcal{O}_{F_{\widetilde{v}}})$ defined in Section 4. Write $L_{\lambda,\tau}$ for the tensor product of the L_{λ_v,τ_v}, regarded as a representation of $G(\mathcal{O}_{F^+,p})$ by letting $G(\mathcal{O}_{F^+,p})$ act on L_{λ_v,τ_v} via $\iota_{\widetilde{v}}$, and for any $\mathcal{O}$-algebra A we write $S_{\lambda,\tau}(U, A)$ for $S(U, L_{\lambda,\tau} \otimes_{\mathcal{O}} A)$.

5.3. **Hecke algebras and Galois representations.** This assumption made above that $\overline{F}^{\ker \mathrm{ad}\, \overline{\rho}|_{G_F}}$ does not contain $F(\zeta_p)$ means that we can and do choose a finite place $v_1 \notin S_p$ of F^+ which splits over F such that v_1 does not split completely in $F(\zeta_p)$, and $\mathrm{ad}\, \overline{\rho}|_{G_F}(\mathrm{Frob}_{v_1}) = 1$. (We make this last assumption in order to simplify the deformation theory at v_1; in particular these assumptions will imply that the local unrestricted lifting ring at v_1 is smooth, and that all liftings are unramified.)

Let $U = \prod_v U_v$ be a compact open subgroup of $G(\mathbb{A}_{F^+}^\infty)$ with U_v a compact open subgroup of $G(F_v^+)$ such that:

- $U_v = G(\mathcal{O}_{F_v^+})$ for all v which split in F other than v_1;
- U_{v_1} is the preimage of the upper triangular matrices under
$$G(\mathcal{O}_{F_{v_1}^+}) \to G(k_{v_1}) \xrightarrow[\iota_{w_1}]{\sim} \mathrm{GL}_n(k_{v_1})$$
where w_1 is a place of F over v_1;
- U_v is a hyperspecial maximal compact subgroup of $G(F_v^+)$ if v is inert in F.

Then U is sufficiently small (by the choice of U_{v_1}). Let $T = S_p \cup \{v_1\}$. We let $\mathbb{T}^{T,\mathrm{univ}}$ be the commutative $\mathcal{O}$-polynomial algebra generated by formal variables $T_w^{(j)}$ for all $1 \le j \le n$, w a place of F lying over a place v of F^+ which splits in F and is not contained in T. The algebra $\mathbb{T}^{T,\mathrm{univ}}$ acts on $S_{\lambda,\tau}(U,\mathcal{O})$ via the Hecke operators
$$T_w^{(j)} := \iota_w^{-1}\left[\mathrm{GL}_n(\mathcal{O}_{F_w})\begin{pmatrix}\varpi_w 1_j & 0\\ 0 & 1_{n-j}\end{pmatrix}\mathrm{GL}_n(\mathcal{O}_{F_w})\right]$$
for $w \notin T$ and ϖ_w a uniformiser in $\mathcal{O}_{F_w}$. We denote by $\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})$ the image of $T^{T,\mathrm{univ}}$ in $\mathrm{End}_{\mathcal{O}}(S_{\lambda,\tau}(U,\mathcal{O}))$.

5.3.1. **Definition.** *We say that a maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{T,\mathrm{univ}}$ with residue field of characteristic p is* automorphic *if for some (λ,τ) as above we have $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \neq 0$. We say that a representation $\overline{\rho} : G_{F^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$ is* automorphic *if there is an automorphic maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{T,\mathrm{univ}}$ such that if $v \notin T$ is a place of F^+ which splits as $v = ww^c$ in F, then $\overline{\rho}|_{G_F}(\mathrm{Frob}_w)$ has characteristic polynomial equal to the image of $X^n + \cdots + (-1)^j(\mathbf{N}w)^{j(j-1)/2}T_w^{(j)}X^{n-j} + \ldots (-1)^n(\mathbf{N}w)^{n(n-1)/2}T_w^{(n)}$.*

In the following we will make a number of arguments that are vacuous unless $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \neq 0$ for the particular (λ,τ) under consideration, but for technical reasons we do not make this assumption. Since $\overline{\rho}$ is a suitable globalization of $\bar{r}$ in the sense of Section 5, it is automorphic by assumption, and there is a maximal ideal $\mathfrak{m}$ of $\mathbb{T}^{T,\mathrm{univ}}$ associated to $\overline{\rho}$ as in Definition 5.3.1. Let $G_{F^+,T} := \mathrm{Gal}(F(T)/F^+)$, where $F(T)$ is the maximal extension of F unramified outside of T and infinity.

5.3.2. **Proposition.** *There is a unique continuous lift*
$$\rho_{\mathfrak{m}} : G_{F^+,T} \to \mathcal{G}_n(\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}})$$
of $\overline{\rho}$, which satisfies

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

(2) $\nu \circ \rho_{\mathfrak{m}} = \varepsilon^{1-n}\delta^n_{F/F^+}$.

(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*
$$X^n + \cdots + (-1)^j(\mathbf{N}w)^{j(j-1)/2}T_w^{(j)}X^{n-j} + \ldots (-1)^n(\mathbf{N}w)^{n(n-1)/2}T_w^{(n)}.$$

(4) *For each place $v \in S_p$, and each $\mathcal{O}$-algebra homomorphism*
$$x : \mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} \to E',$$
where E'/E is a finite extension, the representation $x \circ \rho_{\mathfrak{m}}|_{G_{F_{\tilde{v}}}}$ is potentially crystalline of Hodge type λ_v and inertial type τ_v.

Proof. This may be proved in the same way as Proposition 3.4.4 of [CHT08], making use of Corollaire 5.3 of [Lab09], Theorem 1.1 of [BLGGT12], and Theorem 4.1.5. (Note that the purity assumption in Theorem 4.1.5 holds by Theorem 1.2 of [Car12], which builds on the results of [Shi11] and [CH11].) □

We note that our assumptions on v_1 imply that $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ is free over $\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}[1/p]$ of rank $n!$ (where we adopt the convention that this remark is true if both $\mathbb{T}^T_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}[1/p]$ and $S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ are zero). (This just comes from the fact that the Iwahori invariants of an unramified principal series representation of GL_n have dimension $n!$.)

5.4. **Deformations to $\mathcal{G}_n$.** Let S be a set of places of F^+ which split in F, with $S_p\subseteq S$. As in [CHT08], we will write $F(S)$ for the maximal extension of F unramified outside S and infinity, and from now on we will write $G_{F^+,S}$ for $\mathrm{Gal}(F(S)/F^+)$. Regard $\overline{\rho}$ as a representation of $G_{F^+,S}$. We will freely make use of the terminology (of liftings, framed liftings etc) of Section 2 of [CHT08].

Choose a place $\widetilde{v}_1$ of F above v_1. 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 reduced and p-torsion free quotient of the universal $\mathcal{O}$-lifting ring of $\overline{\rho}|_{G_{F_{\widetilde{v}}}}$. For each $v\in S_p$, write $R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}$ for $R^{\square}_{\overline{\rho}|_{G_{F_{\widetilde{v}}}},\lambda_v,\tau_v}$.

Consider (in the terminology of [CHT08]) the deformation problem

$$\mathcal{S}:=(F/F^+,T,\widetilde{T},\mathcal{O},\overline{\rho},\varepsilon^{1-n}\delta^n_{F/F^+},\{R^{\square}_{\widetilde{v}_1}\}\cup\{R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}\}_{v\in S_p}).$$

There is a corresponding universal deformation $\rho^{\mathrm{univ}}_{\mathcal{S}}:G_{F^+,T}\to\mathcal{G}_n(R^{\mathrm{univ}}_{\mathcal{S}})$ of $\overline{\rho}$. In addition, there is a universal T-framed deformation ring $R^{\square_T}_{\mathcal{S}}$ in the sense of Definition 1.2.1 of [CHT08], which parameterises deformations of $\overline{\rho}$ of type $\mathcal{S}$ together with particular local liftings for each $\widetilde{v}\in\widetilde{T}$. The lifting of Proposition 5.3.2 and the universal property of $\rho^{\mathrm{univ}}_{\mathcal{S}}$ gives an $\mathcal{O}$-homomorphism

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

which is surjective by property (3) of $\rho_{\mathfrak{m}}$.

5.5. **Patching.** We now follow the proof of Theorem 6.8 of [Tho12]. We wish to consider auxiliary sets of primes in order to apply the Taylor–Wiles–Kisin patching method. Since $\overline{\rho}(G_{F(\zeta_p)})$ is adequate by assumption, by Proposition 4.4 of [Tho12] we see that we can (and do) choose an integer $q\geq[F^+:\mathbb{Q}]n(n-1)/2$ and for each $N\geq 1$ a tuple $(Q_N,\widetilde{Q}_N,\{\overline{\psi}_{\widetilde{v}}\}_{v\in Q_N})$ such that

- Q_N is a finite set of finite places of F^+ of cardinality q which is disjoint from T and consists of places which split in F;
- $\widetilde{Q}_N$ consists of a single place $\widetilde{v}$ of F above each place v of Q_N;
- $\mathbf{N}v\equiv 1 \mod p^N$ for $v\in Q_N$;
- for each $v\in Q_N$, $\overline{\rho}|_{G_{F_{\widetilde{v}}}}\cong\overline{s}_{\widetilde{v}}\oplus\overline{\psi}_{\widetilde{v}}$ where $\overline{\psi}_{\widetilde{v}}$ is an eigenspace of Frobenius corresponding to an eigenvalue α_v, on which Frobenius acts semisimply. Write $d^v_N=\dim\overline{\psi}_{\widetilde{v}}$.

For each $v\in Q_N$, let $R^{\overline{\psi}_{\widetilde{v}}}_{\widetilde{v}}$ denote the quotient of $R^{\square}_{\widetilde{v}}$ corresponding to lifts $r:G_{F_{\widetilde{v}}}\to\mathrm{GL}_n(A)$ which are $\ker(\mathrm{GL}_n(A)\to\mathrm{GL}_n(k))$-conjugate to a lift of the form $s\oplus\psi$, where s is an unramified lift of $\overline{s}_{\widetilde{v}}$ and ψ is a lift of $\overline{\psi}_{\widetilde{v}}$ for which the image of inertia is contained in the scalar matrices. We let $\mathcal{S}_{Q_N}$ denote the deformation problem

$$\mathcal{S}_{Q_N} := (F/F^+,\, T \cup Q_N,\, \widetilde{T} \cup \widetilde{Q}_N,\, \mathcal{O},\, \overline{\rho},\, \varepsilon^{1-n}\delta^n_{F/F^+},\\ \{R^{\square}_{\widetilde{v}_1}\} \cup \{R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}\}_{v\in S_p} \cup \{R^{\overline{\psi}_{\widetilde{v}}}_{\widetilde{v}}\}_{v\in Q_N}).$$

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

$$R^{\mathrm{loc}} := \left(\widehat{\otimes}_{v\in S_p} R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}\right) \widehat{\otimes} R^{\square}_{\widetilde{v}_1}$$

where all completed tensor products are taken over $\mathcal{O}$. By Proposition 4.4 of [Tho12], we may also assume that

- the ring $R^{\square_T}_{\mathcal{S}_{Q_N}}$ can be topologically generated over R^{loc} by $q-[F^+:\mathbb{Q}]n(n-1)/2$ elements.

Let $U_0(Q_N) = \prod_v U_0(Q_N)_v$ and $U_1(Q_N) = \prod_v U_1(Q_N)_v$ be the compact open subgroups of $G(\mathbb{A}^\infty_{F^+})$ defined by (for $i = 0,\, 1$) $U_i(Q_N)_v = U_v$ if $v \notin Q_N$, and $U_0(Q_N)_v = \iota^{-1}_{\widetilde{v}}\mathfrak{p}^{\widetilde{v}}_N$, $U_1(Q_N)_v = \iota^{-1}_{\widetilde{v}}\mathfrak{p}^{\widetilde{v}}_{N,1}$ if $v \in Q_N$, where $\mathfrak{p}^{\widetilde{v}}_N$ and $\mathfrak{p}^{\widetilde{v}}_{N,1}$ are the parahoric subgroups defined in [Tho12], corresponding to the partition $n = (n - d^{\widetilde{v}}_N) + d^{\widetilde{v}}_N$. We have natural maps

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

Thus $\mathfrak{m}$ determines maximal ideals of the first three algebras in this sequence which we denote by $\mathfrak{m}_{Q_N}$ for the first two and $\mathfrak{m}$ for the third. Note also that $\mathbb{T}^{T\cup Q_N}_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}} = \mathbb{T}^{T}_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}$ by the proof of Corollary 3.4.5 of [CHT08]. For each $v \in Q_N$ choose an element $\phi_{\widetilde{v}} \in G_{F_{\widetilde{v}}}$ lifting the geometric Frobenius element of $G_{F_{\widetilde{v}}}/I_{F_{\widetilde{v}}}$ and let $\varpi_{\widetilde{v}} \in \mathcal{O}_{F_{\widetilde{v}}}$ be the uniformiser with $\mathrm{Art}_{F_{\widetilde{v}}}\varpi_{\widetilde{v}} = \phi_{\widetilde{v}}|_{F^{\mathrm{ab}}_{\widetilde{v}}}$. As in Proposition 5.9 of [Tho12] there are commuting projection operators $\mathrm{pr}_{\varpi_{\widetilde{v}}} \in \mathrm{End}_{\mathcal{O}}(S_{\lambda,\tau}(U_i(Q_N),\mathcal{O})_{\mathfrak{m}_{Q_N}})$. Write $M = S_{\lambda,\tau}(U,\mathcal{O})_{\mathfrak{m}}$, and for $i = 0,\, 1$ we write

$$M_{i,Q_N} = \left(\prod_{v\in Q_N} \mathrm{pr}_{\varpi_{\widetilde{v}}}\right) S_{\lambda,\tau}(U_i(Q_N),\mathcal{O})_{\mathfrak{m}_{Q_N}}.$$

Let $\mathbb{T}_{i,Q_N}$ denote the image of $\mathbb{T}^{T\cup Q_N}_{\lambda,\tau}(U_i(Q_N),\mathcal{O})$ in $\mathrm{End}_{\mathcal{O}}(M_{i,Q_N})$, let $\Delta_{Q_N} = U_0(Q_N)/U_1(Q_N)$, and let $\mathfrak{a}_{Q_N}$ denote the kernel of the augmentation map

$$\mathcal{O}[\Delta_{Q_N}] \to \mathcal{O}.$$

For $i = 0, 1$ and $\alpha \in F_{\widetilde{v}}$ of non-negative valuation, we have the Hecke operator

$$V_\alpha = \iota^{-1}_{\widetilde{v}}\left(\left[\mathfrak{p}^{\widetilde{v}}_{N,1}\begin{pmatrix} 1_{n-d^{\widetilde{v}}_N} & 0 \\ 0 & A_\alpha \end{pmatrix}\mathfrak{p}^{\widetilde{v}}_{N,1}\right]\right),$$

where $A_\alpha = \mathrm{diag}(\alpha, 1, \ldots, 1)$. Exactly as in the proof of Theorem 6.8 of [Tho12], we have:

(1) The map
$$\big(\prod_{v\in Q_N} \mathrm{pr}_{\varpi_{\widetilde{v}}}\big) : M \to M_{i,Q_N}$$
is an isomorphism.

(2) M_{1,Q_N} is free over $\mathcal{O}[\Delta_{Q_N}]$ with

$$M_{1,Q_N}/\mathfrak{a}_{Q_N} \xrightarrow{\sim} M_{0,Q_N}.$$

(3) For each $v \in Q_N$, there is a character with open kernel $V_{\widetilde{v}} : F_{\widetilde{v}}^\times \to \mathbb{T}_{1,Q_N}^\times$ so that
   (a) for each $\alpha \in F_{\widetilde{v}}$ of non-negative valuation, $V_\alpha = V_{\widetilde{v}}(\alpha)$ on M_{1,Q_N};
   (b) $(\rho_{\mathfrak{m}_{Q_N}} \otimes_{\mathbb{T}^{T\cup Q_N}_{\lambda,\tau}(U_i(Q_N),\mathcal{O})_{\mathfrak{m}_{Q_N}}} \mathbb{T}_{1,Q_N})|_{W_{F_{\widetilde{v}}}} \cong s \oplus \psi$ with s an unramified lift of $\bar{s}_{\widetilde{v}}$ and ψ a lift of $\overline{\psi}_{\widetilde{v}}$ with $I_{F_{\widetilde{v}}}$ acting on ψ via the scalar $(V_{\widetilde{v}} \circ \mathrm{Art}_{F_{\widetilde{v}}}^{-1})$.

The above shows, in particular, that the lift $\rho_{\mathfrak{m}_{Q_N}} \otimes \mathbb{T}_{1,Q_N}$ of $\overline{\rho}$ is of type $\mathcal{S}_{Q_N}$ and gives rise to a surjection $R^{\mathrm{univ}}_{\mathcal{S}_{Q_N}} \twoheadrightarrow \mathbb{T}_{1,Q_N}$. Thinking of Δ_{Q_N} as the product of the inertia subgroups in the maximal abelian p-power order quotient of $\prod_{v\in Q_N} G_{F_{\widetilde{v}}}$, we obtain a homomorphism $\Delta_{Q_N} \to (R^{\mathrm{univ}}_{\mathcal{S}_{Q_N}})^\times$ by considering ψ as above in some basis. We thus have homomorphisms

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

and natural isomorphisms $R^{\mathrm{univ}}_{\mathcal{S}_{Q_N}}/\mathfrak{a}_{Q_N} \cong R^{\mathrm{univ}}_{\mathcal{S}}$ and $R^{\square_T}_{\mathcal{S}_{Q_N}}/\mathfrak{a}_{Q_N} \cong R^{\square_T}_{\mathcal{S}}$, and the surjection $R^{\mathrm{univ}}_{\mathcal{S}_{Q_N}} \twoheadrightarrow \mathbb{T}_{1,Q_N}$ is a homomorphism of $\mathcal{O}[\Delta_{Q_N}]$-algebras.

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

$$0 = L_0 \subset L_1 \subset \cdots \subset L_s = L_{\lambda,\tau}/\pi L_{\lambda,\tau}$$

such that each L_i is $G(\mathcal{O}_{F^+,p})$-stable, and for each $i = 0, 1, \ldots, s-1$, the quotient $\sigma_i := L_{i+1}/L_i$ is absolutely irreducible. This induces a filtration

$$0 = M^0 \subset M^1 \subset \cdots \subset M^s = M/\pi M.$$

We may now patch just as in the proof of [Tho12, Thm. 6.8], keeping track of filtrations as in [Kis09a, § (2.2.9)]. (See also section 4.3 of [GK12] in the case $n = 2$.) More precisely, we patch together the $R^{\mathrm{loc}}[\Delta_{Q_N}]$-modules $M_{1,Q_N} \otimes_{R^{\mathrm{univ}}_{\mathcal{S}_{Q_N}}} R^{\square_T}_{\mathcal{S}_{Q_N}}$ via Lemma 6.10 of [Tho12]. Having made this construction, if we set

$$R_\infty := \left(\widehat{\otimes}_{v\in S_p,\mathcal{O}} R^{\square}_{\widetilde{v}}\right) \widehat{\otimes}_{\mathcal{O}} R^{\square}_{\widetilde{v}_1}[[x_1, \ldots, x_{q-[F^+:\mathbb{Q}]n(n-1)/2}]],$$

$$\begin{aligned}\bar{R}_\infty := \left(\widehat{\otimes}_{v\in S_p,\mathcal{O}} R^{\square,\lambda_v,\tau_v}_{\widetilde{v}}\right) \widehat{\otimes}_{\mathcal{O}} R^{\square}_{\widetilde{v}_1}&[[x_1, \ldots, x_{q-[F^+:\mathbb{Q}]n(n-1)/2}]] \\ &= R^{\mathrm{loc}}[[x_1, \ldots, x_{q-[F^+:\mathbb{Q}]n(n-1)/2}]],\end{aligned}$$

and

$$S_\infty := \mathcal{O}[[z_1, \ldots, z_{n^2\#T}, y_1, \ldots, y_q]],$$

for formal variables $x_1, \ldots, x_{q-[F^+:\mathbb{Q}]n(n-1)/2}$, $y_1, \ldots, y_q$ and $z_1, \ldots, z_{n^2\#T}$, and if we let $\mathfrak{a}$ denote the kernel of the augmentation map $S_\infty \to \mathcal{O}$, then we see that there exists:

- An S_∞-module M_∞ which is simultaneously an $\bar{R}_\infty$-module such that the image of $\bar{R}_\infty$ in $\mathrm{End}(M_\infty)$ is an S_∞-algebra.
- A filtration by $\bar{R}_\infty$-modules

$$0 = M^0_\infty \subset M^1_\infty \subset \cdots \subset M^s_\infty = M_\infty/\pi M_\infty$$

  whose graded pieces are finite free S_∞-modules.
- A surjection of R^{loc}-algebras $\bar{R}_\infty \to R^{\mathrm{univ}}_{\mathcal{S}}$.

- An isomorphism of $\bar{R}_\infty$-modules
$$M_\infty/\mathfrak{a}M_\infty \xrightarrow{\sim} M,$$
which identifies $M^i_\infty/\mathfrak{a}M^i_\infty$ with M^i, and such that the induced $\bar{R}_\infty$-module structure on M coincides with the $R^{\mathrm{univ}}_{\mathcal{S}}$-module structure on M, via the surjection $\bar{R}_\infty \to R^{\mathrm{univ}}_{\mathcal{S}}$ mentioned in the previous point.

We furthermore claim that we can make the above construction so that, for each value of $i = 1, 2, \ldots s$, the (R_∞, S_∞)-bimodule structure on $M^i_\infty/M^{i-1}_\infty$ (arising from its $(\bar{R}_\infty, S_\infty)$-bimodule structure, and the surjection $R_\infty \to \bar{R}_\infty$) and the isomorphism $M^i_\infty/(\mathfrak{a}M^i_\infty, M^{i-1}_\infty) \xrightarrow{\sim} M^i/M^{i-1}$ depends 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.

It will be convenient in the following to refer to a tuple $(a_v)_{\widetilde{v}\in\widetilde{S}_p}$ as a Serre weight, where each a_v is a Serre weight for $\mathrm{GL}_n(k_{\widetilde{v}})$, where $k_{\widetilde{v}}$ is the residue field of $F_{\widetilde{v}}$. We write $F_a = \otimes_{\widetilde{v}\in\widetilde{S}_p} F_{a_v}$, a representation of $G(\mathcal{O}_{F^+,p})$. For a a Serre weight, we denote by M^a_∞ the R_∞/π-module constructed above (using the construction for all (λ, τ)) when $L_i/L_{i-1} \xrightarrow{\sim} F_a$, and we set
$$\mu'_a(\overline{\rho}) = (n!)^{-1} e(M^a_\infty, R_\infty/\pi)$$
and
$$Z'_a(\overline{\rho}) = (n!)^{-1} Z(M^a_\infty).$$
(Note that *a priori* this is only a cycle with $\mathbb{Q}$-coefficients rather than $\mathbb{Z}$-coefficients, but in the situations in which our theorems apply, it will follow that it is in fact a cycle with $\mathbb{Z}$-coefficients.)

For each $v|p$ we write
$$(L_{\lambda_v,\tau_v} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \sum_{a_v} F^{n_{a_v}}_{a_v},$$
so that
$$(L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \sum_{a} F^{n_a}_{a},$$
where $n_a = \prod_v n_{a_v}$.

The following technical, but crucial, lemma is a slight refinement (and generalisation to the n-dimensional setting) of Lemma 4.3.8 of [GK12].

5.5.1. **Lemma.** *For each a, $\mu'_a(\overline{\rho})$ is a non-negative integer. Moreover, for any fixed (λ, τ) the following conditions are equivalent:*

(1) The support of $M \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ 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 $\bar{R}_\infty[1/p]$-module which is everywhere locally free of rank $n!$.

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

(4) $e(\bar{R}_\infty/\pi) = (n!)^{-1} \sum_a n_a e(M^a_\infty, R_\infty/\pi) = \sum_a n_a \mu'_a(\overline{\rho})$.

(5) $Z(\bar{R}_\infty/\pi) = (n!)^{-1} \sum_a n_a Z(M^a_\infty) = \sum_a n_a Z'_a(\overline{\rho})$.

Proof. By Proposition 4.1.2 and our assumptions on the primes in T, $\bar{R}_\infty[1/p]$ is formally smooth, and equidimensional of dimension $q + n^2\#T = \dim S_\infty[1/p]$. Furthermore, the morphism $\operatorname{Spec} \bar{R}_\infty[1/p] \to \operatorname{Spec} R^{\mathrm{loc}}[1/p]$, corresponding to the

$R^{\mathrm{loc}}[1/p]$-algebra structure on $\bar{R}_\infty[1/p]$, induces a bijection on irreducible components; below we denote this bijection by $Z \mapsto Z'$.

Since M_∞ is finite free over S_∞, and the image of $\bar{R}_\infty$ in $\mathrm{End}(M_\infty)$ is an S_∞-algebra, $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ has depth at least $q + n^2\#T$ at every maximal ideal of $\bar{R}_\infty[1/p]$ in its support; so by the Auslander–Buchsbaum formula, it has depth exactly $q + n^2\#T$ at every maximal ideal of $\bar{R}_\infty[1/p]$ in its support, and $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ has projective dimension 0 over $\bar{R}_\infty[1/p]$. Since $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is finite over $S_\infty[1/p]$, we see that it is finite flat over $\bar{R}_\infty[1/p]$, and so its support is a union of irreducible components of $\mathrm{Spec}\, \bar{R}_\infty[1/p]$.

If $Z \subset \mathrm{Spec}\, \bar{R}_\infty[1/p]$ is an irreducible component in the support of $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, then Z is finite over $\mathrm{Spec}\, S_\infty[1/p]$ and of the same dimension, because $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is finite free over $S_\infty[1/p]$. Hence the map $Z \to \mathrm{Spec}\, S_\infty[1/p]$ is surjective. In particular, the fibre of Z over the closed point $\mathfrak{a}$ of $\mathrm{Spec}\, S_\infty[1/p]$ is non-zero, and hence gives a point in the support of $M_\infty/\mathfrak{a}M_\infty = M$ lying in the component Z' of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$ corresponding to Z. As M has rank $n!$ over any point of $R^{\mathrm{univ}}_{\mathcal{S}}[1/p]$ in its support, it follows that $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is in fact locally free of rank $n!$ over all of Z.

Thus we see that $M_\infty \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is faithful over $\bar{R}_\infty[1/p]$, or equivalently, has support consisting of the union of all the irreducible components of $\bar{R}_\infty[1/p]$, if and only if the support of M meets each irreducible component of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$, and that, in this case, it is furthermore locally free of rank $n!$. This shows that (1) and (2) are equivalent.

Using Proposition 1.3.4 of [Kis09a] we also see that each $\mu'_a(\overline{\rho})$ is a non-negative integer, and that

$$e(\bar{R}_\infty/\pi) \geq (n!)^{-1} e(M_\infty/\pi M_\infty, \bar{R}_\infty/\pi) = (n!)^{-1} \sum_a n_a e(M^a_\infty, R_\infty/\pi)$$

with equality if and only if M_∞ is a faithful $\bar{R}_\infty$-module (in which case, as already observed, it is necessarily locally free of rank $n!$). Thus (2) and (4) are equivalent.

If M_∞ is a faithful $\bar{R}_\infty$-module, then $\bar{R}_\infty$, being a subring of the ring of S_∞-endomorphisms of the finite S_∞-module M_∞, is finite over S_∞, and so $R^{\mathrm{univ}}_{\mathcal{S}}$, which is a quotient of $\bar{R}_\infty/\mathfrak{a}$, is a finite $\mathcal{O}$-module. This shows that (2) implies (3). Now assume (3), so that $R^{\mathrm{univ}}_{\mathcal{S}}$ is a finite $\mathcal{O}$-algebra; we see by Proposition 1.5.1 of [BLGGT10] that the image of

$$\mathrm{Spec}\, R^{\mathrm{univ}}_{\mathcal{S}} \to \mathrm{Spec}\, R^{\mathrm{loc}}$$

meets every component of $\mathrm{Spec}\, R^{\mathrm{loc}}[1/p]$. Hence, since $M \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a faithful $R^{\mathrm{univ}}_{\mathcal{S}}$-module by assumption, we see that (1) holds. Thus (1)–(4) are all equivalent.

It remains to show that (5) is equivalent to the other conditions. From Lemma 2.2.8, we see that (5) implies (4). Now assume that (2) holds. Then M_∞ is π-torsion free and generically free of rank $n!$ over each component of $\mathrm{Spec}\, \bar{R}_\infty$, so by Proposition 2.2.13 we have

$$Z(\bar{R}_\infty/\pi) = (n!)^{-1} Z(M_\infty/\pi M_\infty).$$

Identifying each of the modules $M_\infty/\pi M_\infty$, $M^i_\infty/M^{i-1}_\infty$, etc., with the corresponding sheaves on $\operatorname{Spec} R_\infty/\pi$ that they give rise to, we see from Lemma 2.2.7 that we have

$$\begin{aligned} Z(M_\infty/\pi M_\infty) &= \sum_i Z(M^i_\infty/M^{i-1}_\infty) \\ &= \sum_a n_a Z(M^a_\infty). \end{aligned}$$

The result follows. □

For each Serre weight $a \in (\mathbb{Z}^n_+)^{\operatorname{Hom}(k,\mathbb{F})}$, let λ_a be a fixed lift of a. As in Remark 4.1.7(1), we may inductively define unique integers $\mu_a(\bar{r})$ such that for any Serre weight b, if we write

$$(L_{\lambda_b,\mathbb{1}} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_a F_a^{m_{a,b}},$$

then

$$e(R^{\square}_{\bar{r},\lambda_b,\mathbb{1}}/\pi) = \sum_a m_{a,b}\mu_a(\bar{r}).$$

(In Remark 4.1.7(1) we were assuming the Breuil–Mézard conjecture, but the inductive process defining the $\mu_a(\bar{r})$ does not rely on any cases of the conjecture.) Similarly, we may inductively define unique cycles $\mathcal{C}_a$ such that

$$Z(R^{\square}_{\bar{r},\lambda_b,\mathbb{1}}/\pi) = \sum_a m_{a,b}\mathcal{C}_a.$$

We now prove the main result of this section, a conditional equivalence between the geometric formulation of the Breuil–Mézard conjecture, the numerical formulation, and a global statement. For the convenience of the reader, we recall the various assumptions of this section in the statement of the theorem.

5.5.2. **Theorem.** *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be a finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_n(\mathbb{F})$ be a continuous representation, with $\mathbb{F}$ a finite extension of $\mathbb{F}_p$. Assume that there is a suitable globalization $\overline{\rho}$ of $\bar{r}$ as in Section 5.1 (for example, by Corollary A.7 this is guaranteed if $p \nmid n$ and Conjecture A.3 holds for $\bar{r}$).*

Suppose that the equivalent conditions of Lemma 5.5.1 hold whenever each $\lambda_v = \lambda_{a_v}$ for some Serre weight a_v and each $\tau_v = \mathbb{1}$.

Then, if $\lambda \in (\mathbb{Z}^n_+)^{\operatorname{Hom}_{\mathbb{Q}_p}(K,E)}$ and τ is an inertial type for G_K, and if we write

$$(L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_a F_a^{n_a},$$

the following conditions are equivalent:

(1) *The equivalent conditions of Lemma 5.5.1 hold when $\lambda_v = \lambda$ and $\tau_v = \tau$ for each $v|p$.*
(2) $e(R^{\square}_{\bar{r},\lambda,\tau}/\pi) = \sum_a n_a\mu_a(\bar{r})$.
(3) $Z(R^{\square}_{\bar{r},\lambda,\tau}/\pi) = \sum_a n_a\mathcal{C}_a$.

Proof. Consider conditions (4) and (5) of Lemma 5.5.1 in the case that each $\lambda_v = \lambda_{b_v}$ for some Serre weight b_v and each $\tau_v = \mathbb{1}$. Note that in this case we have $(L_{\lambda,\tau} \otimes_{\mathcal{O}} \mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_a F_a^{m_{a,b}}$ where $m_{a,b} = \prod_v m_{a_v,b_v}$. Throughout this proof we will write $\overline{\rho}_v$ for $\overline{\rho}|_{G_{F_{\widetilde{v}}}}$ (which is isomorphic to $\bar{r}$).

As remarked above, the conditions on v_1 ensure that $R^{\square}_{\widetilde{v}_1}$ is formally smooth over $\mathcal{O}$, say $R^{\square}_{\widetilde{v}_1} \cong \mathcal{O}[[t_1,\dots,t_{n^2}]]$. Our assumption that the equivalent conditions of Lemma 5.5.1 hold implies that

$$\begin{aligned}\sum_a m_{a,b}\mu'_a(\overline{\rho}) &= e(\bar{R}_\infty/\pi)\\ &= \prod_{v|p} e(R^{\square}_{\overline{\rho}_v,\lambda_{b_v},\mathbb{1}}/\pi)\\ &= \prod_{v|p}\sum_{a_v} m_{a_v,b_v}\mu_{a_v}(\bar{r})\\ &= \sum_a m_{a,b}\prod_{v|p}\mu_{a_v}(\bar{r}),\end{aligned}$$

where the second equality holds by Proposition 1.3.8 of [Kis09a]. Similarly

$$\begin{aligned}\sum_a m_{a,b}Z'_a(\overline{\rho}) &= Z(\bar{R}_\infty/\pi)\\ &= \prod_{v|p} Z(R^{\square}_{\overline{\rho}_v,\lambda_{b_v},\mathbb{1}}/\pi)\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]])\\ &= \prod_{v|p}\sum_{a_v} m_{a_v,b_v}\mathcal{C}_{a_v}\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]])\\ &= \sum_a m_{a,b}\prod_{v|p}\mathcal{C}_{a_v}\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]]),\end{aligned}$$

where we have made use of Lemma 2.2.14. If we define a partial order on the tuples $(a_v)_{v|p}$ of Serre weights by writing $(a_v)_{v|p}\ge(b_v)_{v|p}$ to mean that each $a_v\ge b_v$, then an easy induction using Lemma 4.1.1 shows that we must in fact have that for each Serre weight a,

$$\mu'_a(\overline{\rho}) = \prod_{v|p}\mu_{a_v}(\bar{r})$$

and

$$Z'_a(\overline{\rho}) = \prod_{v|p}\mathcal{C}_{a_v}\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]]).$$

We now consider the conditions of Lemma 5.5.1 in the case that each $\lambda_v=\lambda$ and each $\tau_v=\tau$. By definition (and Lemma 2.2.14), we have

$$Z(\bar{R}_\infty/\pi) = \prod_{v|p} Z(R^{\square}_{\overline{\rho}_v,\lambda,\tau}/\pi)\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]]).$$

Using the relation above, we have

$$\begin{aligned}\sum_a n_a Z'_a(\overline{\rho}) &= \sum_a\prod_{v|p} n_{a_v}\mathcal{C}_{a_v}\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]])\\ &= \prod_{v|p}\sum_{a_v} n_{a_v}\mathcal{C}_{a_v}\times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]]).\end{aligned}$$

Now, condition (5) of Lemma 5.5.1 states that

$$Z(\bar{R}_\infty/\pi) = \sum_a n_a Z'_a(\overline{\rho}),$$

which is equivalent to

$$\prod_{v|p} Z(R^{\square}_{\bar{\rho}_v,\lambda,\tau}/\pi) \times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]])$$
$$= \prod_{v|p}\sum_{a_v} n_{a_v}\mathcal{C}_{a_v} \times Z(\mathbb{F}[[x_1,\dots,x_{q-[F^+:\mathbb{Q}]n(n-1)/2},t_1,\dots,t_{n^2}]])$$

and thus to

$$Z(R^{\square}_{\bar{r},\lambda,\tau}/\pi) = \sum_a n_a\mathcal{C}_a,$$

giving the equivalence of (1) and (3). The equivalence of (1) and (2) may be proved by a formally identical argument. $\square$

5.5.3. **Remark.** We regard the assumption of Theorem 5.5.2 (that the conditions of Lemma 5.5.1 hold for all Serre weights) as a strong form of the weight part of Serre's conjecture, saying that each set of components of the local crystalline deformation rings in low weight is realised by a global automorphic Galois representation. Under this assumption, Theorem 5.5.2 may be regarded as saying that instances of the Breuil–Mézard conjecture (and its geometric refinement) are equivalent to certain $R = \mathbb{T}$ theorems.

One case in which Theorem 5.5.2 has unconditional consequences is the case that $n = 2$ and $\lambda = 0$, where the automorphy lifting theorems of [Kis09b] and [Gee06] can be applied. We now give such an application, where we make use of the results of [GK12] to give a particularly clean statement. Note that when $n = 2$, the inductive definition of the cycles $\mathcal{C}_a$ above is trivial, and we simply set $\mathcal{C}_a := Z(R^{\square}_{\bar{r},\lambda_a,\mathbb{1}}/\pi)$.

Suppose that $K/\mathbb{Q}_p$ is absolutely unramified, and that $n = 2$; then we say that a Serre weight a is a *predicted Serre weight for* $\bar{r}$ if it is one of the weights predicted in [BDJ10]. We say that a is *regular* if we have $0 \le a_{\sigma,1} - a_{\sigma,2} < p-1$ for each $\sigma \in \mathrm{Hom}(k,\mathbb{F})$.

5.5.4. **Theorem.** *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be an absolutely unramified finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_2(\mathbb{F})$ be a continuous representation, with $\mathbb{F}$ a finite extension of $\mathbb{F}_p$. Suppose that every predicted Serre weight for $\bar{r}$ is regular. Then, for any inertial type τ for G_K, if we write $(L_{0,\tau}\otimes_{\mathcal{O}}\mathbb{F})^{\mathrm{ss}} \xrightarrow{\sim} \oplus_a F_a^{n_a}$, we have $Z(R^{\square}_{\bar{r},0,\tau}/\pi) = \sum_a n_a\mathcal{C}_a$.*

Proof. This will follow from Theorem 5.5.2 once we have checked that all of the hypotheses hold. Since $n = 2$ and $p > 2$, we have $p \nmid n$ and thus a suitable globalization will exist provided that Conjecture A.3 holds for $\bar{r}$; but this follows from the proof of Theorem A.1.2 of [GK12] (which shows that $\bar{r}$ has a potentially Barsotti–Tate lift) and Lemma 4.4.1 of *op.cit.* (which shows that any potentially Barsotti–Tate representation is potentially diagonalizable).

It remains to check that the equivalent conditions of Lemma 5.5.1 hold whenever $\lambda_v = 0$ and $\tau_v = \tau$ (for an arbitrary τ) for all $v|p$, and whenever τ_v is trivial for all v and each λ_v corresponds to a Serre weight. In the first (respectively second) case, this follows from Lemma 4.4.1 (respectively Lemma 4.4.2) of [GK12], exactly as in the proof of Corollary 4.4.3 of [GK12]. $\square$

5.5.5. **Remark.** In particular, Theorem 5.5.4 confirms Conjecture 1.4 of [BM12] in the case that each $k_\tau = 2$, and extends Théorème 1.5 of *ibid.* to arbitrary types.

Appendix A. Realising local representations globally

In this appendix we realise local representations globally, using potential automorphy theorems. In the case $n = 2$, analogous results were proved by similar techniques in Appendix A of [GK12]. We would like to thank Robert Guralnick and Florian Herzig for the proof of the following fact.

A.1. **Lemma.** *Suppose that $p > 2$ and $p \nmid n$. Let $(\mathrm{GL}_n.2)(\mathbb{F}_{p^m})$ denote the subgroup of $\mathrm{GL}_{2n}(\mathbb{F}_{p^m})$ generated by block diagonal matrices of the form $(g, {}^t g^{-1})$ and a matrix J with $J(g, {}^t g^{-1})J^{-1} = ({}^t g^{-1}, g)$. Then for m sufficiently large, both $(\mathrm{GL}_n.2)(\mathbb{F}_{p^m}) \subset \mathrm{GL}_{2n}(\overline{\mathbb{F}}_p)$ and $\mathrm{GL}_n(\mathbb{F}_{p^m}) \subset \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ are adequate.*

Proof. In the case of $\mathrm{GL}_n(\mathbb{F}_{p^m})$, this is immediate from Theorem 1.2 and Lemma 1.4 of [Gur12b]. For $(\mathrm{GL}_n.2)(\mathbb{F}_{p^m})$, since we are assuming that $(p, 2n) = 1$, it follows from Theorem 1.5 of [Gur12a] (which shows the corresponding result for the underlying algebraic group $\mathrm{GL}_n.2$) and the proof of Theorem 1.2 of [Gur12b] that the result will be true provided that $\mathrm{Ext}^1_{(\mathrm{GL}_n.2)(\mathbb{F}_{p^m})}(V, V) = 0$ for m sufficiently large, where $V = \mathbb{F}_{p^m}^{2n}$ is the space on which $(\mathrm{GL}_n.2)(\mathbb{F}_{p^m})$ acts.

The result then follows easily from inflation-restriction, the corresponding result for $\mathrm{GL}_n(\mathbb{F}_{p^m})$ (which is proved in the course of the proof of Theorem 1.2 of [Gur12b]), and the fact that the standard representation of $\mathrm{GL}_n(\mathbb{F}_{p^m})$ has a different central character to its dual provided that $p^m > 3$. $\square$

A.2. **Proposition.** *Let $K/\mathbb{Q}_p$ be a finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ be a continuous representation, with $p \nmid n$. Then there is a CM field L with maximal totally real subfield L^+, and a continuous irreducible representation $\overline{\rho} : G_{L^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$, such that*

- *each place $v|p$ of L^+ splits in L;*
- *for each place $v|p$ of L^+, $L^+_v \cong K$ and there is a place $\widetilde{v}$ of L lying over v such that $\overline{\rho}|_{G_{L_{\widetilde{v}}}} \cong \bar{r}$;*
- *$\nu \circ \overline{\rho} = \overline{\varepsilon}^{1-n} \delta^n_{L/L^+}$, where δ_{L/L^+} is the quadratic character corresponding to L/L^+;*
- *$\overline{\rho}^{-1}(\mathrm{GL}_n(\overline{\mathbb{F}}_p) \times \mathrm{GL}_1(\overline{\mathbb{F}}_p)) = G_L$;*
- *$\overline{\rho}(G_{L(\zeta_p)}) = \mathrm{GL}_n(\mathbb{F}_{p^m})$ for some sufficiently large m as in Lemma* A.1 *(so in particular, $\overline{\rho}(G_{L(\zeta_p)})$ is adequate);*
- *$\overline{L}^{\ker \overline{\rho}}$ does not contain $L(\zeta_p)$;*
- *if $v \nmid p$ is a finite place of L^+, then $\overline{\rho}|_{G_{L^+_v}}$ is unramified.*

Proof. Choose m such that $\bar{r}(G_K) \subset \mathrm{GL}_n(\mathbb{F}_{p^m})$ and m is sufficiently large as in Lemma A.1, and set $\mathbb{F} = \mathbb{F}_{p^m}$. We now apply Proposition 3.2 of [Cal12]; in the notation of that result, we take $G = \mathcal{G}_n(\mathbb{F})$, and we let E be any totally real field with the property that if $v|p$ is a place of E, then $E_v \cong K$. We let $w \nmid p$ be a finite place of E, and we choose a character $\overline{\chi}_w : G_{E_w} \to \mathbb{F}^\times$ such that $\overline{\chi}_w|_{I_{F_w}}$ surjects onto $\mathbb{F}^\times$. We take $F = E(\zeta_p)$, and we take S to be the set of places of E which either divide p or else are infinite, together with the place w. We take $c_v = j \in \mathcal{G}_n(\mathbb{F})$ for each infinite place v of S, and for each place $v|p$ of E we let H_v/E_v correspond to $\overline{K}^{\ker \bar{r}}/K$, and let $\phi_v : \mathrm{Gal}(H_v/E_v) \to \mathcal{G}_n(\mathbb{F})$ correspond to $(\bar{r}, \overline{\varepsilon}^{1-n}) : G_K \to \mathrm{GL}_n(\mathbb{F}) \times \mathrm{GL}_1(\mathbb{F}) \subset \mathcal{G}_n(\mathbb{F})$. We let $H_w = \overline{E}_w^{\ker \overline{\chi}_w}$, and let $\phi_w = (1, \overline{\chi}_w) : \mathrm{Gal}(H_w/E_w) \to \mathrm{GL}_n(\mathbb{F}) \times \mathrm{GL}_1(\mathbb{F}) \subset \mathcal{G}_n(\mathbb{F})$.

Writing M^+/E for the field denoted K in [Cal12], we see that M^+ is a totally real field, and that we have a surjective representation $\overline{\rho} : G_{M^+} \to \mathcal{G}_n(\mathbb{F})$, with the properties that

- for each place $v|p$ of M^+, $M^+_v \cong K$ and there is a place $\widetilde{v}$ of L lying over v such that $\overline{\rho}|_{M_{\widetilde{v}}} \cong \bar{r}$,
- w splits completely in M^+, and if $w'|w$ then $\nu \circ \overline{\rho}|_{G_{M^+_{w'}}} = \overline{\chi}$,
- for each place $v|\infty$ of M^+, $\overline{\rho}(c_v) = j$, where c_v is a complex conjugation at v,
- $\nu \circ \overline{\rho}|_{G_{M^+_v}} = \overline{\varepsilon}^{1-n}$ if $v|p$,
- $\overline{M^+}^{\ker \overline{\rho}}$ does not contain $M^+(\zeta_p)$.

Define M/M^+ by $G_M = \overline{\rho}^{-1}(\mathrm{GL}_n(\mathbb{F}) \times \mathrm{GL}_1(\mathbb{F}))$. Since $\overline{\rho}(c_v) = j$ for each complex conjugation c_v, we see that M is an imaginary CM field. Similarly, we see that each place of M^+ lying over p or w splits in M.

Now consider the character $\overline{\phi} := (\nu \circ \overline{\rho})\overline{\varepsilon}^{n-1}\delta^n_{M/M^+}$. By construction, we see that $\overline{\phi}(c_v) = 1$ for each complex conjugation c_v, that $\overline{\phi}|_{G_{M^+_v}} = 1$ for each place $v|p$, and that $\overline{\phi}|_{I_{M^+_{w'}}} = \overline{\chi}|_{I_{M^+_{w'}}}$ for each place $w'|w$. Replace M^+ by $N^+ := \overline{M^+}^{\ker \overline{\phi}}$, M by $N := N^+M$, and $\overline{\rho}$ by $\overline{\rho}|_{G_{N^+}}$. Then N^+/M^+ is a totally real abelian extension in which all places dividing p split completely, and it is linearly disjoint from $M^+(\zeta_p)$ over M^+ (by considering the ramification at places above w).

To complete the proof, it suffices to choose a totally real finite Galois extension L^+/N^+ such that

- each place of N^+ lying over p splits completely in L^+,
- $\overline{\rho}|_{G_{L^+_v}}$ is unramified for all finite places $v \nmid p$ of L^+, and
- L^+ is linearly disjoint from $\overline{N^+}^{\ker \overline{\rho}|_{G_{N^+}}}(\zeta_p)$ over N^+.

The first two of these conditions concern the splitting behaviour of a finite number of places of N^+, so the existence of an extension satisfying all but the third condition is guaranteed by Lemma 2.2 of [Tay03]. In order to satisfy the third condition as well, for each of the (finitely many) Galois intermediate fields $\overline{N^+}^{\ker \overline{\rho}|_{G_{N^+}}}(\zeta_p)/L_i/N^+$ with the property that $\mathrm{Gal}(L_i/N^+)$ is simple one chooses a finite place of N^+ which does not split completely in L_i, and arranges that it splits completely in L^+ (cf. the proof of Proposition 6.2 of [BLGHT11]). □

In order to study the Breuil–Mézard conjecture globally, we need to be able to realise local mod p representations as part of global automorphic Galois representations. In order to do this, we will apply the above construction, and then show that the representation $\overline{\rho}$ is potentially automorphic. This essentially follows from the results of [BLGGT10]. However, the main theorems of [BLGGT10] make no attempt to control the local behaviour at finite places of the extensions over which potential automorphy is proved, while it is crucial for us that we do not change the local fields at places dividing p. Fortunately, a simple trick using restriction of scalars that we learned from Richard Taylor allows us to deduce the potential automorphy results that we need from those of [BLGGT10]. We will need the following conjecture.

A.3. **Conjecture.** *Let $K/\mathbb{Q}_p$ be a finite extension, and $\bar{r} : G_K \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ a continuous representation. Then $\bar{r}$ has a potentially diagonalizable lift $r : G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ which has regular Hodge–Tate weights.*

A.4. **Remark.** If $n = 2$ then Conjecture A.3 is easily verified by a Galois cohomology calculation (for example, any potentially Barsotti–Tate lift is potentially diagonalizable, and the result is then immediate in the irreducible case, and in the reducible case is a special case of Lemma 6.1.6 of [BLGG12]). We anticipate that a similar argument works more generally.

A.5. **Lemma.** *Let $p \nmid 2n$ be prime. Let $K/\mathbb{Q}_p$ be a finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ be a continuous representation. Assume that Conjecture* A.3 *holds for $\bar{r}$. Let $\overline{\rho} : G_{L^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$ be the representation provided by Lemma* A.2. *Then there is a lift $\rho : G_{L^+} \to \mathcal{G}_n(\overline{\mathbb{Q}}_p)$ of $\overline{\rho}$ such that*

- *$\nu \circ \rho = \varepsilon^{1-n}\delta^n_{L/L^+}$,*
- *ρ is unramified outside of places dividing p, and*
- *if $w|p$ is a place of L, then $\rho|_{G_{F_w}}$ is potentially diagonalizable with regular Hodge–Tate weights.*

Proof. In the case that $p > 2(n + 1)$, this follows at once from Theorem 4.3.1 of [BLGGT10], together with Conjecture A.3. As explained in the first paragraph of Theorem A.4.1 of [BLGG13], the only obstacle to extending Theorem 4.3.1 of [BLGGT10] to the case $p \nmid 2n$ is the problem of relaxing Proposition 3.3.1 of [BLGGT10] to hold under this assumption, and in particular, it is necessary to know that the induction of a representation from a degree 2 subgroup is adequate. However, in our situation, this is guaranteed by Lemma A.1. $\square$

At this point, we would like to apply Theorem 4.5.1 of [BLGGT10] to establish that for some CM extension F/L, $\rho|_{G_F}$ is (in the terminology of [BLGGT10]) automorphic. However, for our applications we need the places of L above p to split completely in F, which is not guaranteed by Theorem 4.5.1 of [BLGGT10]. Accordingly, we need to give a slight modification of the proof of *loc. cit.*

A.6. **Proposition.** *Maintain the notation and assumptions of Lemma* A.5. *Then there is a CM extension F/L which is linearly disjoint from $L^{\ker\overline{\rho}}(\zeta_p)$ over L, such that each place of L lying over p splits completely in F, and $\rho|_{G_F}$ is automorphic in the terminology of* [BLGGT10].

Proof. The proof is a straightforward modification of the arguments of [BLGGT10], specifically Theorem 3.1.2 and Proposition 3.3.1 of *op. cit.* We sketch the details. We will freely use the notation and terminology of [BLGGT10]. As in the proof of Proposition 3.3.1 of [BLGGT10], we may choose a character $\psi : G_L \to \overline{\mathbb{Q}}_p^\times$ such that

- ψ is crystalline at all places above p.
- $R := I(\rho|_{G_{L^+}} \otimes (\psi, \varepsilon^{-n}\delta^{n+1}_{L/L^+})) : G_{L^+} \to \mathrm{GSp}_{2n}(\overline{\mathbb{Q}}_p)$ has multiplier ε^{1-2n}, and is crystalline with distinct Hodge–Tate weights at all places lying over p.
- $\bar{R}(G_{L^+(\zeta_p)})$ is adequate. (This follows from the choice of $\mathbb{F} = \mathbb{F}_{p^m}$, which was chosen large enough that the conclusion of Lemma A.1 holds.)

We now employ a slight variant of the proof of Theorem 3.1.2 of [BLGGT10]. We let the set $\mathcal{I}$ of *loc. cit.* consist of just the single element $\{1\}$, we set $n_1 = 2n$, and we put $l_1 = p$, $r_1 := \bar{R}$, $F = F_0 = L^+$, $F^{(\mathrm{avoid})} = L(\zeta_p)$. We then apply the constructions of *loc. cit.*, choosing in particular an auxiliary positive integer N, and constructing a geometrically irreducible scheme $\widetilde{T} = T_{\bar{r}_1 \times \bar{r}_1'}$ over $\operatorname{Spec} L^+(\zeta_N)^+$.

Choose a solvable extension $(L')^+/L^+$ of totally real fields such that

- $(L')^+$ is linearly disjoint from $\overline{L}^{\ker \overline{\rho}}(\zeta_p)$ over L^+, and
- for each place $v|p$ of $(L')^+(\zeta_N)^+$, there is a point $P_v \in \widetilde{T}((L')^+(\zeta_N)^+_v)$ with $v(t_1(P_v)) < 0$.

We now construct a geometrically irreducible scheme $\widetilde{T}'$ over $\operatorname{Spec}(L')^+(\zeta_N)^+$ in exactly the same way as $\widetilde{T}$ is constructed over $\operatorname{Spec} L^+(\zeta_N)^+$, and we then set $T' := \operatorname{Res}_{(L')^+(\zeta_N)^+/L^+}\widetilde{T}'$, a geometrically irreducible scheme over $\operatorname{Spec} L^+$.

We can then apply Proposition 3.1.1 of [BLGGT10], to find a finite Galois extension F^+/L^+ of totally real fields in which the places of L^+ over p split completely, and a point of $P \in T'(F^+)$ which satisfies the assumptions (relating to linear disjointness from a certain field, and on the $v(T_i(P)))$ on the field F' in the proof of Theorem 3.1.2 of [BLGGT10]. (We can assume that the places of L^+ over p split completely in F^+ by the assumption on the places of $(L')^+(\zeta_N)^+$ lying over p.)

Regarding P as an $F^+(L')^+(\zeta_N)^+$-point of $\widetilde{T}'$, it then follows exactly as in *loc. cit.* that there is a continuous lift $R' : G_{F^+(L')^+(\zeta_N)^+} \to \mathrm{GSp}_{2n}(\overline{\mathbb{Q}}_p)$ of the restriction $\overline{\rho}|_{G_{F^+(L')^+(\zeta_N)^+}}$ such that

- R' is ordinary, and
- R' is automorphic.

Let M be a quadratic totally imaginary extension of $F^+(L')^+(\zeta_N)^+$ such that

- all places of $F^+(L')^+(\zeta_N)^+$ above p split in M, and
- M is linearly disjoint from $\overline{L}^{\ker \overline{\rho}} F^+(L')^+(\zeta_p)$ over $F^+(L')^+$.

It follows from Theorem 4.2.1 of [BLGGT10] that $R|_{G_M}$ is automorphic, and then from Lemma 2.1.1 of [BLGGT10] that $\rho|_{G_{F^+(L')^+(\zeta_N)^+L}}$ is automorphic. Since the extension $F^+(L')^+(\zeta_N)^+L/F^+L$ is solvable, it follows from Lemma 1.4 of [BLGHT11] that $\rho|_{G_{F^+L}}$ is automorphic. Since the places of L^+ over p split completely in F^+, the claim follows upon taking $F = F^+L$. $\square$

A.7. **Corollary.** *Suppose that $p \nmid 2n$, that $K/\mathbb{Q}_p$ is a finite extension, and let $\bar{r} : G_K \to \mathrm{GL}_n(\overline{\mathbb{F}}_p)$ be a continuous representation for which Conjecture* A.3 *holds. Then there is an imaginary CM field F and a continuous irreducible representation $\overline{\rho} : G_{F^+} \to \mathcal{G}_n(\overline{\mathbb{F}}_p)$ such that*

- *each place $v|p$ of F^+ splits in F, and has $F_v^+ \cong K$,*
- *for each place $v|p$ of F^+, there is a place $\widetilde{v}$ of F lying over v with $\overline{\rho}|_{G_{F_{\widetilde{v}}}}$ isomorphic to $\bar{r}$,*
- *$\overline{\rho}$ is unramified outside of p,*
- *$\overline{\rho}^{-1}(\mathrm{GL}_n(\overline{\mathbb{F}}_p) \times \mathrm{GL}_1(\overline{\mathbb{F}}_p)) = G_F$,*
- *$\overline{\rho}(G_{F(\zeta_p)})$ is adequate,*
- *$\overline{F}^{\ker \overline{\rho}}$ does not contain $F(\zeta_p)$,*
- *$\overline{\rho}$ is automorphic in the sense of Definition* 5.3.1*, and in particular $\mu \circ \overline{\rho} = \bar{\varepsilon}^{1-n}\delta_{F/F^+}^n$.*

Proof. This follows from Proposition A.6 and the theory of base change between GL_n and unitary groups, cf. Proposition 2.2.7 of [Ger09]. $\square$

References

[BC09] Joël Bellaïche and Gaëtan Chenevier, *Families of Galois representations and Selmer groups*, Astérisque (2009), no. 324, xii+314.

[BDJ10] Kevin Buzzard, Fred Diamond, and Frazer Jarvis, *On Serre's conjecture for mod l Galois representations over totally real fields*, Duke Math. J. **155** (2010), no. 1, 105–161.

[BLGG11] Tom Barnet-Lamb, Toby Gee, and David Geraghty, *The Sato-Tate Conjecture for Hilbert Modular Forms*, J. Amer. Math. Soc. **24** (2011), no. 2, 411–469.

[BLGG12] Thomas Barnet-Lamb, Toby Gee, and David Geraghty, *Congruences between Hilbert modular forms: constructing ordinary lifts*, Duke Mathematical Journal **161** (2012), no. 8, 15211580.

[BLGG13] ______, *Serre weights for rank two unitary groups*, Math. Ann. (to appear) (2013).

[BLGGT10] Tom Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Potential automorphy and change of weight*, preprint, 2010.

[BLGGT12] ______, *Local-global compatibility for $l = p$, II*, Annales scientifiques de l'ENS (to appear) (2012).

[BLGHT11] Tom Barnet-Lamb, David Geraghty, Michael Harris, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy II*, Publ. Res. Inst. Math. Sci. **47** (2011), no. 1, 29–98.

[BM02] Christophe Breuil and Ariane 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** (2002), no. 2, 205–310, With an appendix by Guy Henniart.

[BM12] ______, *Multiplicités modulaires raffinées*, Bull. Soc. Math. de France. (to appear) (2012).

[Cal12] Frank Calegari, *Even Galois representations and the Fontaine–Mazur conjecture. II*, J. Amer. Math. Soc. **25** (2012), no. 2, 533–554.

[Car12] Ana Caraiani, *Local-global compatibility and the action of monodromy on nearby cycles*, Duke Mathematical Journal **161** (2012), no. 12, 23112413.

[CDT99] Brian Conrad, Fred Diamond, and Richard Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. **12** (1999), no. 2, 521–567.

[CH11] Gaëtan Chenevier and Michael Harris, *Construction of automorphic Galois representations, II*, preprint available at `http://www.math.polytechnique.fr/~chenevier/pub.html`, 2011.

[CHT08] Laurent Clozel, Michael Harris, and Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Pub. Math. IHES **108** (2008), 1–181.

[Con10] William Conley, *Inertial types and automorphic representations with prescribed ramification*, Ph.D. thesis, UCLA, 2010.

[Con12] ______, *In preparation*, 2012.

[Gee06] Toby Gee, *A modularity lifting theorem for weight two Hilbert modular forms*, Math. Res. Lett. **13** (2006), no. 5-6, 805–811.

[Gee11] ______, *Automorphic lifts of prescribed types*, Mathematische Annalen **350** (2011), 107–144, 10.1007/s00208-010-0545-z.

[Ger09] David Geraghty, *Modularity lifting theorems for ordinary Galois representations*, preprint, 2009.

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

[Gur12a] Robert Guralnick, *Adequate subgroups II*, Bulletin of Mathematical Sciences **2** (2012), no. 1, 193–203.

[Gur12b] Robert M. Guralnick, *Adequacy of representations of finite groups of Lie type (Appendix A to Dieulefait's paper "Automorphy of* $\mathrm{Symm}^5(\mathrm{GL}(2))$ *and base change")*, 2012.

[Her09] Florian Herzig, *The weight in a Serre-type conjecture for tame n-dimensional Galois representations*, Duke Math. J. **149** (2009), no. 1, 37–116.

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

[Hum06] James E. Humphreys, *Modular representations of finite groups of Lie type*, London Mathematical Society Lecture Note Series, vol. 326, Cambridge University Press, Cambridge, 2006.

[Jan03] Jens Carsten Jantzen, *Representations of algebraic groups*, second ed., Mathematical Surveys and Monographs, vol. 107, American Mathematical Society, Providence, RI, 2003.

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

[Kis09a] ———, *The Fontaine-Mazur conjecture for* GL_2, J. Amer. Math. Soc. **22** (2009), no. 3, 641–690.

[Kis09b] ———, *Moduli of finite flat group schemes, and modularity*, Annals of Math.(2) **170** (2009), no. 3, 1085–1180.

[Kis10] ———, *The structure of potentially semi-stable deformation rings*, Proceedings of the International Congress of Mathematicians, vol. II (2010), 294–311.

[Lab09] Jean-Pierre Labesse, *Changement de base CM et séries discrètes*, preprint, 2009.

[Mat89] Hideyuki Matsumura, *Commutative ring theory*, second ed., Cambridge Studies in Advanced Mathematics, vol. 8, Cambridge University Press, Cambridge, 1989, Translated from the Japanese by M. Reid.

[Paš05] Vytautas Paškūnas, *Unicity of types for supercuspidal representations of* GL_N, Proc. London Math. Soc. (3) **91** (2005), no. 3, 623–654.

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

[Sav05] David Savitt, *On a conjecture of Conrad, Diamond, and Taylor*, Duke Math. J. **128** (2005), no. 1, 141–197.

[Shi11] Sug Woo Shin, *Galois representations arising from some compact Shimura varieties*, Ann. of Math. (2) **173** (2011), no. 3, 1645–1741.

[Sno09] Andrew Snowden, *On two dimensional weight two odd representations of totally real fields*, 2009, preprint.

[Tay03] Richard Taylor, *On icosahedral Artin representations. II*, Amer. J. Math. **125** (2003), no. 3, 549–566.

[Tho12] Jack Thorne, *On the automorphy of l-adic Galois representations with small residual image*, Journal of the Institute of Mathematics of Jussieu (2012), 1–66.

Mathematics Department, Northwestern University, 2033 Sheridan Rd., Evanston, IL 60208

E-mail address, Matthew Emerton: `emerton@math.northwestern.edu`

E-mail address, Toby Gee: `gee@math.northwestern.edu`