

# MINIMAL MODULARITY LIFTING FOR NONREGULAR SYMPLECTIC REPRESENTATIONS

---

FRANK CALEGARI and DAVID GERAGHTY

Appendix by FRANK CALEGARI, DAVID GERAGHTY, and MICHAEL HARRIS

## Abstract

*We prove a minimal modularity lifting theorem for Galois representations (conjecturally) associated to Siegel modular forms of genus 2 which are holomorphic limits of discrete series at infinity.*

## Contents

1. Introduction . . . . .	1
2. Notation . . . . .	6
3. Some commutative algebra . . . . .	11
4. Deformations of Galois representations . . . . .	12
5. Siegel 3-folds . . . . .	20
6. Galois representations associated to modular forms . . . . .	35
7. Properties of cohomology groups . . . . .	46
8. $q$ -expansions of Siegel modular forms . . . . .	56
9. Modularity lifting . . . . .	78
Appendix. Bloch–Kato conjectures for automorphic motives . . . . .	80
References . . . . .	91

## 1. Introduction

In this article, we prove a minimal modularity lifting theorem for Galois representations (conjecturally) associated to Siegel modular forms  $\pi$  for the group  $\mathrm{GSp}(4)/\mathbf{Q}$  such that  $\pi_\infty$  is a holomorphic limit of discrete series. An example of what we can prove with these methods is the following.

DUKE MATHEMATICAL JOURNAL

Advance publication—final volume, issue, and page numbers to be assigned.

© 2019 DOI [10.1215/00127094-2019-0044](https://doi.org/10.1215/00127094-2019-0044)

Received 22 March 2017. Revision received 8 May 2019.

2010 *Mathematics Subject Classification*. Primary 11F80; Secondary 11F33, 11F46.

## THEOREM 1.1

Let  $r : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$  be a continuous irreducible representation satisfying the following conditions:

- (1) The restriction  $r|_{G_{\mathbf{Q}_p}}$  is ordinary with Hodge–Tate weights  $[0, 0, j-1, j-1]$  for some integer  $j$  satisfying  $p-1 > j \geq 4$ .
- (2) If  $\alpha$  and  $\beta$  are the unit root eigenvalues of Frobenius on  $D_{\mathrm{cris}}(V)$ , then

$$(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2\beta^2 - 1) \not\equiv 0 \pmod{p}.$$

- (3) The image of  $\bar{r}|_{G_{\mathbf{Q}(\zeta_p)}}$  contains  $\mathrm{Sp}_4(\mathbf{F}_p)$ .
- (4) For a prime  $x \neq p$ , the image of inertia at  $x$  is unipotent, and the image of any generator of tame inertia has the same number of Jordan blocks mod  $p$  as it does in characteristic 0.
- (5) The representation  $\bar{r}$  is modular of level  $N(\bar{r})$  and weight  $(j, 2)$ .

Then  $r$  is modular, that is, there exists a cuspidal Siegel modular Hecke eigenform  $F$  of weight  $(j, 2)$  such that

$$L(r, s) = L(F, s),$$

where  $L(F, s)$  is the spinor  $L$ -function of  $F$ .

We deduce Theorem 1.1 from our main result, which we now state. (We refer to Sections 4 and 6.4 for precise details concerning ramification behavior, level subgroups, and the exact definition of minimal deformations.) Let  $\epsilon$  denote the  $p$ -adic cyclotomic character. Let  $\mathcal{O}$  be the ring of integers of a finite extension  $K$  of  $\mathbf{Q}_p$ . Let  $\bar{r} : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(k)$  be a continuous irreducible representation whose similitude character  $\nu(\bar{r})$  on inertia at  $p$  is the mod- $p$  reduction of  $\epsilon^{1-j}$ . Suppose that  $\bar{r}|_{G_{\mathbf{Q}_p}}$  contains an unramified subspace of dimension 2 on which  $\mathrm{Frob}_p$  acts by the scalars  $\alpha$  and  $\beta$ , respectively, where

$$(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2\beta^2 - 1) \neq 0.$$

Suppose further that  $\bar{r}$  has big image (explicitly, satisfies Assumption 4.1) and that  $\bar{r}|_{G_{\mathbf{Q}_x}}$  for a prime  $x \neq p$  is either unramified or is one of the types listed in Assumption 4.3. Let  $Y_1(N)$  denote the (open) Siegel modular variety of level  $N = N(\bar{r})$  over  $\mathrm{Spec}(\mathcal{O})$ , where  $N$  is determined by  $\bar{r}$  as in Section 5, and let  $\omega(j, 2)$  denote the coherent sheaf on  $Y_1(N)$  whose complex sections define Siegel modular forms of weight  $(j, 2)$  for some integer  $p-1 > j \geq 4$ . Let  $\mathbf{T}$  denote the subring of endomorphisms of

$$eH^0(Y_1(N), \omega(j, 2) \otimes K/\mathcal{O}) \simeq \varinjlim eH^0(Y_1(N), \omega(j, 2) \otimes \mathcal{O}/\varpi^n)$$

(where  $e = e_{\alpha, \beta}$  is a certain ordinary projection; see Section 6.4) generated by Hecke operators at primes not dividing  $Np$ . Let  $R^{\min}$  denote the universal minimal deformation ring of  $\bar{r}$  (see Definition 4.6 for more details).

**THEOREM 1.2**

*Suppose that there exist both a maximal ideal  $\mathfrak{m}$  of  $\mathbf{T}$  and a corresponding representation  $\bar{r}_{\mathfrak{m}} : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(k)$  which is isomorphic to  $\bar{r}$ . Let  $R^{\min}$  denote the universal minimal ordinary deformation ring of  $\bar{r}$ . Suppose that  $p - 1 > j \geq 4$ . Then there is an isomorphism  $R^{\min} \xrightarrow{\sim} \mathbf{T}_{\mathfrak{m}}$ , and moreover, the  $\mathbf{T}_{\mathfrak{m}}$ -module  $eH^0(Y_1(N), \omega(j, 2) \otimes K/\mathcal{O})_{\mathfrak{m}}^{\vee}$  is free as a  $\mathbf{T}_{\mathfrak{m}}$ -module.*

The proof follows the strategy of [16]. The main ingredients are showing that there exists a map from  $R^{\min}$  to  $\mathbf{T}_{\mathfrak{m}}$  (see Theorem 6.17) and proving that the cohomology of the subcanonical extension  $\omega(j, 2)^{\mathrm{sub}}$  of  $\omega(j, 2)$  to a smooth toroidal compactification  $X_1(N)$  of  $Y_1(N)$  vanishes outside degrees 0 and 1 (see Theorem 5.1—in the case of classical modular curves this step was trivial).

As an application of the results of the present paper, we establish in Theorem A.1 of the Appendix some cases of the Bloch–Kato conjecture for adjoint motives of modular abelian surfaces.

*1.1. Comparison with previous methods*

The first modularity theorems which applied to nonregular motives were the results of Buzzard and Taylor (see [14], [15]) on 2-dimensional odd Artin representations  $V$ . The idea of these papers can roughly be described as follows. Using known cases of Serre’s conjecture, one deduces that  $\bar{\rho}$  is modular, where  $\rho$  is the representation associated to some  $p$ -adic realization of  $V$  for some  $p$ . Using modularity theorems in *regular* weight, one then proves that a big Hecke algebra is modular. Specializing to weight 1, one deduces the existence of an *overconvergent* eigenform  $f$  corresponding to  $V$ . Under a nondegeneracy assumption on  $V$  ( $\bar{\rho}$  is  $p$ -distinguished), one constructs (using companion forms) a second Hida family which specializes to a *second* eigenform  $g$ . Using the geometric properties of  $U$ , one shows that  $f$  and  $g$  converge deeply into the supersingular locus. The Fourier coefficients  $a_n$  of  $f$  and  $g$  for  $(n, p) = 1$  are determined by  $V$ . One then constructs a suitable linear combination  $h = (\alpha f - \beta g)/(\alpha - \beta)$  which converges over the entire modular curve, and is thus classical by rigid GAGA. The formal rigid geometry employed by these papers has been generalized by various authors, in particular by Kassaei [44]. One may well ask whether this approach can be applied to Siegel modular forms of weight  $(2, 2)$ —work of Tilouine and his collaborators has made great progress in this direction. The modularity lifting result for (regular weight) Hida families has been established in many

cases by Genestier and Tilouine [31] (see also Pilloni [56]). Significant progress has also been made in the theory of canonical subgroups and the geometry of Siegel modular varieties. One difficulty, however, is that the Fourier expansions of Siegel modular forms are *not* determined by the Hecke eigenvalues. This is a difficulty which must be overcome in such an approach. (Various classicality results for overconvergent forms can be established without using  $q$ -expansions—see, for example, [44] and [59]—but these results only apply to forms of sufficiently noncritical slope.) The difficulty of dealing with  $q$ -expansions manifests itself for our approach also—we are forced to prove (“by hand”) various properties of Fourier expansions of Hecke eigenforms in Section 8.2.

### 1.2. Abelian surfaces

It would be desirable to weaken the assumption  $j \geq 4$  in Theorem 1.2 to  $j \geq 2$ , since the case  $j = 2$  includes the representations associated to the Tate modules of abelian surfaces. The only point in our arguments in which we use the fact that  $j \geq 4$  is to deduce that  $H^2(X_1(N), \omega(j, 2)^{\text{sub}}) = 0$  for the subcanonical extension  $\omega(j, 2)^{\text{sub}}$  of  $\omega(j, 2)$  to a smooth toroidal compactification  $X_1(N)$  of  $Y_1(N)$ . If this vanishing holds for  $j = 2$ , then our theorem would also apply to these cases. On the other hand, one does *not* expect vanishing here, because one expects that singular Siegel modular forms should contribute cohomology in these degrees. However, we need only the weaker result that the image of  $H^2(X_1(N), \omega(j, 2)^{\text{can}})$  in  $H^2(X_1(N), \omega(j, 2)^{\text{sub}})$  is zero after localization at a sufficiently non-Eisenstein maximal ideal  $\mathfrak{m}$ . We expect this to always be true for  $j = 2$ , although we were not able to prove it. On the other hand, using the ideas of Khare and Thorne in [46], one can dispense with proving this under the very strong supplementary hypothesis that there exists a *characteristic 0* form of weight  $N = N(\bar{F})$  which gives rise to  $\bar{r}$ . In particular, by using the arguments of the proof of [46, Theorem 6.29], one should be able to prove the analogue of Theorem 1.1 in weight  $(2, 2)$  assuming the existence of an auxiliary Siegel modular form  $G$  of the same level also of weight  $(2, 2)$  with  $\bar{r}_G = \bar{r}$ .

### 1.3. Recent developments

As of January 2019, there have been a number of developments related to the main theme of this paper, in particular in the preprints [13] and [58], the latter of which establishes the potential modularity of abelian surfaces over totally real fields. The introduction to [13] explains a number of innovations which made those results possible, so we confine ourselves here to only a few salient remarks. The first is that the vanishing conjecture for  $H^2$  localized at  $\mathfrak{m}$  mentioned in Section 1.2 remains unresolved, and the methods of [13] blend the techniques of this paper (and [16]) with arguments from [58]. A second point is that the paper [58] develops a concep-

tual method to define (normalized) Hecke operators at  $p$ , and in particular establishes the action of these operators on higher cohomology (which is essential for the main results of [58] and [13]). In this paper, it suffices to construct the action of these Hecke operators on  $H^0$  which is significantly easier. The methods we use in Section 8.4 to do this are admittedly disagreeable, relying as they do on arguments using  $q$ -expansions. Thus the reader is encouraged to consult [58, Section 7] and [13, Section 4.5] for a more geometric construction of these operators. An analysis of the normalization factors for Hecke operators required in [58] also sheds some light on another phenomenological feature of this paper which readers may find surprising. On the Galois side, there is essentially no difference (in the ordinary setting) between working in the (irregular) weight  $(j, 2)$  for  $j > 2$  and working in the (irregular) weight  $(2, 2)$ . On the other hand, the Hecke operators at  $p$  (particularly  $T_{p,2}$ ) behave quite differently in weight  $(2, 2)$ . In our context, this arises most noticeably in Section 8.5 (via Lemma 8.12), which one should compare to [58, Section 11.1] (warning: the convention of that paper is that Pilloni's  $T_{p,1}$  is our  $T_{p,2}$  and vice versa, and the spherical version of the operator  $T$  in [58] is equal in weight  $(2, 2)$  up to translation by a multiple of  $T_{p,0}$  to the operator we call  $Q_2$ ). Lastly, the paper [13] develops a geometric version of the doubling argument (see [13, Section 5]). This provides a much more robust explanation (in a slightly different setting) for what in this paper occupies most of Section 8 and consists of a sequence of tricky and not entirely intuitive series of manipulations with  $q$ -expansions. (Note that the geometric doubling argument of [13] is only written for weight  $(2, 2)$  but the method applies in principle to the weights  $(j, 2)$  which we consider in this paper.) Finally, the very observant reader will notice that the doubling argument of [13] applies in weight  $(2, 2)$  to the space of ordinary forms at Klingen level, whereas in this paper we essentially prove (in the same weight) a tripling result at spherical level. Neither of these results immediately imply the other. The “extra” copy of the space of forms can be interpreted as giving rise to a space of *nonordinary* forms of weight  $(p + 1, p + 1)$ . See Remark 8.18 for further discussion on this point, which we also discuss in a different context below.

It is natural to ask whether one should expect any genuine difficulties in modifying the geometric doubling argument of [13] to the setting of this paper. We now offer some speculative remarks to address this point (using notation from [13]). Let  $\pi_p$  be a smooth admissible irreducible unramified representation of  $\mathrm{GL}_2(\mathbf{Q}_p)$  (over  $\mathbf{C}$ ) which is not trivial. (For example,  $\pi_p$  could be the local constituent of an automorphic representation  $\pi$  associated to a classical modular form.) Let  $\mathrm{Sph} = \mathrm{GL}_2(\mathbf{Z}_p)$ , and let  $\mathrm{Iw}$  denote the Iwahori subgroup of  $\mathrm{Sph}$ . The classical theory of oldforms is a reflection of the fact that

$$\dim \pi_p^{\mathrm{Iw}} = 2 = 2 \cdot \dim \pi_p^{\mathrm{Sph}}$$

and the characteristic 0 version of doubling is the statement that the span of the spherical vector  $v$  under the operator  $U_p$  is all of  $\pi_p^{\text{Iw}}$ . The integral version of this statement is false in general. For example, given a classical ordinary modular eigenform  $f$  of weight  $k \geq 2$ , the span of  $f \bmod p$  under  $U_p$  is simply  $f$  because  $T_p = U_p \bmod p$  in these weights. However, some version of this result *does* hold in weight  $k = 1$ , and it is this property which is leveraged to prove local-global compatibility results in [16]. Let us now replace  $\text{GL}_2(\mathbf{Q}_p)$  by  $\text{GSp}_4(\mathbf{Q}_p)$ , and let  $\text{Kli}$  and  $\text{Iw}$  denote the Klingen and Iwahori subgroups, respectively, of  $\text{Sph} = \text{GSp}_4(\mathbf{Z}_p)$  (denoted elsewhere in this paper by  $\Pi$  and  $I$ , respectively). Now (for the  $\Pi_p$  of interest) we will have

$$\dim \Pi_p^{\text{Iw}} = 8 = 2 \cdot 4 = 2 \dim \Pi_p^{\text{Kli}} = 8 \dim \Pi_p^{\text{Sph}}.$$

The factor 8 here may be interpreted as the order of the Weyl group of  $\text{GSp}(4)$ . More prosaically, the oldforms in  $\Pi_p^{\text{Iw}}$  correspond to a choice of eigenvalues  $\alpha$  and  $\alpha\beta$  for the Hecke operators  $U_{\text{Iw}(p),1}$  and  $U_{\text{Iw}(p),2}$ , respectively, whereas the oldforms in  $\Pi_p^{\text{Kli}}$  correspond to a choice of eigenvalues  $\alpha + \beta$  and  $\alpha\beta$  for the Hecke operators  $U_{\text{Kli}(p),1}$  and  $U_{\text{Kli}(p),2} = U_{\text{Iw}(p),2}$ . When one passes from  $\pi_p^{\text{Sph}}$  to  $\pi_p^{\text{Iw}}$  for weight 1 modular forms or from  $\Pi_p^{\text{Kli}}$  to  $\Pi_p^{\text{Iw}}$  for weight (2, 2) Siegel modular forms, the property of being ordinary turns out to be automatically preserved on the corresponding space of oldforms. However, this is not a priori true when passing from  $\Pi_p^{\text{Sph}}$  to  $\Pi_p^{\text{Iw}}$ , and so one would have to see in any geometric version of this argument a way of dealing with the nonordinary forms.

#### 1.4. Results of Arthur

In Section 7.2, we make use of the results of [3], which sketches how the results of [4] on orthogonal and symplectic groups can be extended to the general symplectic group  $\text{GSp}_4$ . At the time of the initial submission of this paper, these results of Arthur are conditional on the stabilization of the twisted trace formula. (We direct the reader to [30] for the most up-to-date status of these results for  $\text{GSp}_4$ .)

## 2. Notation

We fix a prime  $p$  and let  $\mathcal{O}$  be the ring of integers of a finite extension  $K$  of  $\mathbf{Q}_p$  with residue field  $k$ . We let  $\mathcal{C}_{\mathcal{O}}$  denote the category of complete local Noetherian  $\mathcal{O}$ -algebras  $R$  with residue field isomorphic to  $k$  (via the structural homomorphism  $\mathcal{O} \rightarrow R$ ).

We let

$$\epsilon : G_{\mathbf{Q}} \rightarrow \mathbf{Z}_p^{\times}$$

denote the cyclotomic character. The Hodge–Tate weight of  $\epsilon \mid G_{\mathbf{Q}_p}$  is  $-1$ .

If  $L$  is a finite extension of  $\mathbf{Q}_l$  for some prime  $l$ , then we let  $\text{Art}_L : L^{\times} \rightarrow W_L^{\text{ab}}$  denote the Artin map, normalized so that uniformizers correspond to geometric Frobe-

nus elements. If  $\gamma$  is an element of some ring  $R$ , then we define the character

$$\lambda(\gamma) : G_L \longrightarrow R^\times$$

to be the unramified character which takes the geometric Frobenius element  $\text{Frob}_L$  to  $\gamma$ , when this character is well defined.

### 2.1. The group $\text{GSp}_4$

Let  $G = \text{GSp}_4 = \{M \in \text{GL}_4 : M^t J M = \nu \cdot J \text{ for some } \nu \in \text{GL}_1\}$ , where

$$J := \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.$$

The group  $\text{Sp}_4$  is the subgroup consisting of elements with  $\nu = 1$ . We let  $B \subset G$  be the Borel subgroup consisting of upper triangular matrices. The Lie algebras of  $G$  and  $B$  are denoted  $\mathfrak{g}$  and  $\mathfrak{b}$ , while those of  $\text{Sp}_4$  and  $B \cap \text{Sp}_4$  are denoted  $\mathfrak{g}^0$  and  $\mathfrak{b}^0$ . Let  $P \subset G$  denote the Siegel parabolic, that is, the stabilizer of the plane spanned by the first two standard basis vectors. Let  $\Pi \subset G$  denote the Klingen parabolic, which is the stabilizer of the line spanned by the first standard basis vector. We denote the Levi subgroup of  $P$  (resp.,  $\Pi$ ) by  $M = M_P$  (resp.,  $M_\Pi$ ). We have  $M \cong \text{GL}_2 \times \text{GL}_1$ .

Let  $T$  denote the diagonal torus in  $\text{GSp}_4$ , and let  $X^*(T)$  denote its character group. We identify  $X^*(T)$  with the lattice  $\mathbf{Z}^3$  by associating to  $(a, b; c)$  the character

$$\text{diag}(t_1, t_2, \nu t_2^{-1}, \nu t_1^{-1}) \mapsto t_1^a t_2^b \nu^c.$$

We identify the cocharacter group  $X_*(T)$  with  $\mathbf{Z}^3$  by associating the triple  $(\alpha, \beta; \gamma)$  with the cocharacter

$$t \mapsto \text{diag}(t^\alpha, t^\beta, t^{\gamma-\beta}, t^{\gamma-\alpha}).$$

The natural pairing on  $X^*(T) \times X_*(T)$  is then  $\langle (a, b; c), (\alpha, \beta; \gamma) \rangle \mapsto a\alpha + b\beta + c\gamma$ .

The positive roots of  $G$  with respect to the Borel  $B$  are given by  $\alpha_1 := (1, -1; 0)$ ,  $\alpha_2 := (0, 2; -1)$ ,  $\alpha_3 = (1, 1; -1)$ , and  $\alpha_4 = (2, 0; -1)$ . Of these,  $\alpha_1$  and  $\alpha_2$  are the simple roots. We let  $\rho = (2, 1; -3/2)$  denote the half-sum of the positive roots. The coroots are  $\alpha_1^\vee = (1, -1; 0)$ ,  $\alpha_2^\vee = (0, 1; 0)$ ,  $\alpha_3^\vee = (1, 1; 0)$ , and  $\alpha_4^\vee = (1, 0; 0)$ . The intersection  $B \cap M$  is a Borel subgroup of  $M$ . The corresponding positive root is  $\alpha_1$ .

#### Definition 2.1

We define the set  $X^*(T)_G^+$  to be the set  $\{\lambda \in X^*(T) : \langle \lambda, \alpha_i^\vee \rangle \geq 0 \forall i\}$  of weights which are dominant with respect to  $B$ . Explicitly,

$$X^*(T)_G^+ = \{(a, b; c) \in X^*(T) : a \geq b \geq 0\}.$$

Similarly, we define the set of weights  $X^*(T)_M^+ := \{(a, b; c) \in X^*(T) : \langle \lambda, \alpha_1^\vee \rangle \geq 0\}$  which are dominant with respect to  $B \cap M$ . Explicitly, this is

$$X^*(T)_M^+ = \{(a, b; c) \in X^*(T) : a \geq b\}.$$

Note that the natural action of  $M$  on the plane spanned by the first two (resp., the last two) standard basis vectors is the irreducible representation of highest weight  $(1, 0; 0)$  (resp.,  $(0, -1; 1)$ ).

We let  $W_G = N_G(T)/T$  denote the Weyl group of  $G$ , and we define  $W_M$  and  $W_{M_\Gamma}$  similarly. Let  $s_0, s_1$  denote the generators for the Weyl group  $W_G$  given in [41, Section 2]. We fix a set of Kostant representatives  $W^M = \{\tilde{w}_0, \tilde{w}_1, \tilde{w}_2, \tilde{w}_3\}$  for  $W_M \setminus W_G$  by setting  $\tilde{w}_0 = 1$ ,  $\tilde{w}_1 = s_1$ ,  $\tilde{w}_2 = s_1 s_0$ , and  $\tilde{w}_3 = s_1 s_0 s_1$ . Note that each  $\tilde{w}_i$  has length  $i$ . We let  $w \in W_G$  act on  $X^*(T)$  by  $(w\lambda)(t) = \lambda(w^{-1}tw)$ . Then we have

$$\begin{aligned}\tilde{w}_1(a, b; c) &= (a, -b; b + c), \\ \tilde{w}_2(a, b; c) &= (b, -a; a + c), \\ \tilde{w}_3(a, b; c) &= (-b, -a; a + b + c).\end{aligned}$$

The longest element of  $W_G$  which we denote by  $w_0$  acts via  $w_0(a, b; c) = (b, a; c)$ .

Note that the collection of representatives  $W^M$  is precisely the set of  $w \in W_G$  such that  $w(X^*(T)_G^+) \subset X^*(T)_M^+$ . We let  $C_0 \subset X^*(T)_\mathbf{R} := X^*(T) \otimes_{\mathbf{Z}} \mathbf{R}$  denote the closed dominant Weyl chamber. In other words,  $C_0 = \{(a, b; c) \in \mathbf{R}^3 : a \geq b \geq 0\}$ . For  $i = 1, 2, 3$ , we define the chambers  $C_i := \tilde{w}_i(C_0)$ .

## 2.2. The group $\mathrm{GSp}_4(\mathbf{R})$

Let

$$h : \mathrm{Res}_{\mathbf{C}/\mathbf{R}}(\mathbf{G}_m)(\mathbf{R}) = \mathbf{C}^\times \rightarrow G(\mathbf{R}) = \mathrm{GSp}_4(\mathbf{R})$$

be the homomorphism sending  $x + iy$  to the matrix

$$\begin{pmatrix} xI_2 & yS \\ -yS & xI_2 \end{pmatrix},$$

where

$$S := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let  $K^h$  denote the centralizer of  $h$  in  $G(\mathbf{R})$  (acting by conjugation). Then since  $h(i) = J$ , we see that  $K^h = \mathbf{R}^\times K_\infty$ , where  $K_\infty$  is the maximal compact subgroup of  $G(\mathbf{R})$  given by the fixed points of the Cartan involution  $g \mapsto (g^t)^{-1}$ . The similitude



character restricts to a surjective map  $\nu : K_\infty \rightarrow \{\pm 1\}$  and the kernel  $K_{\infty,1}$  is the connected component of the identity. Then we have explicitly,

$$K_{1,\infty} = \left\{ \begin{pmatrix} SAS & SB \\ -BS & A \end{pmatrix} \in G(\mathbf{R}) : A^t A + B^t B = I_2, A^t B = B^t A \right\}.$$

The map

$$\begin{aligned} K_{\infty,1} &\longrightarrow \mathrm{GL}_2(\mathbf{C}), \\ \begin{pmatrix} SAS & SB \\ -BS & A \end{pmatrix} &\longmapsto A + iB \end{aligned}$$

induces an isomorphism between  $K_{\infty,1}$  and  $U(2)$ . We let  $H_1 \subset K_{\infty,1}$  denote the preimage of the diagonal compact torus in  $U(2)$  and let  $H := \mathbf{R}_{>0}^\times H_1 \subset K^h$ . Let  $\mathfrak{h} = \mathrm{Lie} H$ , let  $\mathfrak{k}^h = \mathrm{Lie} K^h$ , and so on. Then we have

$$\mathfrak{h} = \left\{ h(t_1, t_2; z) := \begin{pmatrix} z & 0 & 0 & t_1 \\ 0 & z & t_2 & 0 \\ 0 & -t_2 & z & 0 \\ -t_1 & 0 & 0 & z \end{pmatrix} : t_1, t_2, z \in \mathbf{R} \right\}.$$

We use subscripts to denote complexifications of Lie algebras and Lie groups; thus  $H_{\mathbf{C}}$  and  $\mathfrak{h}_{\mathbf{C}}$  denote the complexifications of  $H$  and  $\mathfrak{h}$ . Then  $\mathfrak{h}_{\mathbf{C}} = \mathrm{Lie} H_{\mathbf{C}} = \{h(t_1, t_2; t) : t_1, t_2, z \in \mathbf{C}\}$  and the surjective map  $\exp : \mathfrak{h}_{\mathbf{C}} \rightarrow H_{\mathbf{C}}$  sends  $h(t_1, t_2; z)$  to

$$\exp(z) \begin{pmatrix} \cos t_1 & 0 & 0 & \sin t_1 \\ 0 & \cos t_2 & \sin t_2 & 0 \\ 0 & -\sin t_2 & \cos t_2 & 0 \\ -\sin t_1 & 0 & 0 & \cos t_1 \end{pmatrix}.$$

Thus its kernel is  $\{h(t_1, t_2; z) : t_1, t_2 \in 2\pi\mathbf{Z}, z \in 2\pi i\mathbf{Z}\}$ . We define the lattice  $X^*(H_{\mathbf{C}}) \subset \mathfrak{h}_{\mathbf{C}}^*$  to be the subspace consisting of differentials of (complex analytic) characters of  $H_{\mathbf{C}}$ . Equivalently,  $X^*(H_{\mathbf{C}})$  is the subset of  $X^*(\mathbf{C}^\times \times H_{1,\mathbf{C}}) = \{\lambda \in \mathfrak{h}_{\mathbf{C}}^* : \lambda(\ker(\exp : \mathfrak{h}_{\mathbf{C}} \rightarrow H_{\mathbf{C}})) \subset 2\pi i\mathbf{Z}\}$  consisting of differentials of characters of  $\mathbf{C}^\times \times K_{1,\mathbf{C}}$  which factor through the multiplication map  $\mathbf{C}^\times \times H_{1,\mathbf{C}} \rightarrow H_{\mathbf{C}}$ . We fix an isomorphism

$$\{(a, b; c) \in \mathbf{Z}^3 : a + b \equiv c \pmod{2}\} \xrightarrow{\sim} X^*(H_{\mathbf{C}})$$

by letting  $(a, b; c)$  correspond to the linear form

$$h(t_1, t_2; z) \mapsto at_1i + bt_2i + cz$$

on  $\mathfrak{h}_{\mathbf{C}}$ . This extends by linearity to an isomorphism  $\mathbf{C}^3 \rightarrow \mathfrak{h}_{\mathbf{C}}^*$ .

Let  $V^\pm \subset \mathbf{C}^4$  be the subspace where  $h(i)$  acts via  $\pm i$ . Then each  $V^\pm$  is isotropic and we have an orthogonal direct sum  $\mathbf{C}^4 = V^- \oplus V^+$ . Let  $Q^- \subset G(\mathbf{C})$  denote the stabilizer of  $V^-$ . Consider the Hodge decomposition

$$\mathfrak{g}_{\mathbf{C}} = \mathfrak{g}^{0,0} \oplus \mathfrak{g}^{-1,1} \oplus \mathfrak{g}^{1,-1},$$

where  $\mathfrak{g}^{p,q}$  is the subspace on which  $h(z)$  acts via  $z^{-p}\bar{z}^{-q}$ . Then we have  $\mathfrak{g}^{0,0} = \mathfrak{k}_{\mathbf{C}}^h$  and we let  $\mathfrak{p}^+ = \mathfrak{g}^{-1,1}$ ,  $\mathfrak{p}^- = \mathfrak{g}^{1,-1}$ . We also let  $P^\pm$  denote the subgroup of  $G(\mathbf{C})$  generated by  $\exp(\mathfrak{p}^\pm)$ . Then we have

$$Q^- = K_{\mathbf{C}}^h P^- \quad \text{and} \quad \text{Lie } Q^- = \mathfrak{k}_{\mathbf{C}}^h \oplus \mathfrak{p}^-.$$

Moreover,  $K_{\mathbf{C}}^h$  is the Levi component of  $Q^-$  and  $P^-$  is its unipotent radical. Let

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -i \end{pmatrix}, \quad f_2 = \begin{pmatrix} 0 \\ 1 \\ -i \\ 0 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 0 \\ -i \\ 1 \\ 0 \end{pmatrix}, \quad f_4 = \begin{pmatrix} -i \\ 0 \\ 0 \\ 1 \end{pmatrix} \in \mathbf{C}^4.$$

Then  $f_1, f_2$  are a basis of  $V^-$  and  $f_3, f_4$  are a basis of  $V^+$ . With respect to the basis  $f_1, \dots, f_4$  of  $\mathbf{C}^4$ , an element

$$k = \begin{pmatrix} SAS & SB \\ -BS & A \end{pmatrix} \in K_{1,\infty}$$

acts via

$$C^{-1}kC = \begin{pmatrix} SAS - iSBS & 0 \\ 0 & A + iB \end{pmatrix} \quad \text{where } C := \begin{pmatrix} I_2 & -iS \\ -iS & I_2 \end{pmatrix}.$$

Note that the Cayley transform  $C$  conjugates the Siegel parabolic  $P(\mathbf{C})$  to  $Q^-$ . Let  $\Phi \subset X^*(H_{\mathbf{C}})$  denote the root system defined by the adjoint action of  $H_{\mathbf{C}}$  on  $\mathfrak{g}_{\mathbf{C}}$ . The compact roots  $\Phi_c$  are those appearing in  $\mathfrak{k}_{\mathbf{C}}^h$ , while the noncompact roots  $\Phi_n$  are those appearing in  $\mathfrak{p}^+ \oplus \mathfrak{p}^-$ . We choose a system of positive roots  $\Phi^+$  in such a way that the set of positive noncompact roots  $\Phi_n^+ = \Phi^+ \cap \Phi_n$  coincides with the roots in  $\mathfrak{p}^+$ . (We do this in order to be consistent with the conventions of [9, Section 2.4].) We are then forced to take  $\Phi^+$  to be the set of roots appearing in  $C(\text{Lie } \bar{B})C^{-1}$ , where  $\bar{B} \subset G$  is the Borel subgroup of lower triangular matrices. With respect to the identification of  $X^*(H_{\mathbf{C}})$  as a subset of  $\mathbf{Z}^3$  given above, we then have

$$\begin{aligned} \Phi_c^+ &= \{(1, -1; 0)\}, \\ \Phi_n^+ &= \{(0, 2; 0), (1, 1; 0), (2, 0; 0)\}. \end{aligned}$$

This can be seen easily from the fact that  $C^{-1}h(t_1, t_2; 0)C = \text{diag}(-it_1, -it_2, it_2, it_1)$ .

*Definition 2.2*

We let  $X^*(H_{\mathbf{C}})_{K_{\mathbf{C}}^h}^+$  denote the set of weights which are dominant with respect to the system of positive roots  $\Phi_c^+$ . In other words,  $X^*(H_{\mathbf{C}})_{K_{\mathbf{C}}^h}^+ = \{(a, b; c) \in X^*(H_{\mathbf{C}}) : a \geq b\}$ .

This set parameterizes the irreducible complex analytic representations of  $K_{\mathbf{C}}^h$ . For  $\mu \in X^*(H_{\mathbf{C}})_{K_{\mathbf{C}}^h}^+$ , we let  $V_{\mu}$  denote the corresponding irreducible representation of highest weight  $\mu$ .

We note that the natural representation of  $K_{\mathbf{C}}^h$  on  $V^-$  (resp.,  $V^+$ ) is the irreducible representation of highest weight  $(0, -1; 1)$  (resp.,  $(1, 0; 1)$ ). Note also that the similitude character  $\nu : H_{\mathbf{C}} \rightarrow \mathbf{C}^{\times}$  has weight  $(0, 0; 2)$ .

**3. Some commutative algebra**

We recall here some formalism from [16] for proving modularity lifting results in contexts where the Hecke algebra has “codimension 1” over the ring of diamond operators. The notion of “balanced” below plays the role of “codimension 1” for the nonregular group rings  $S_N := \mathcal{O}[(\mathbf{Z}/p^N \mathbf{Z})^q]$ .

*3.1. Balanced modules*

Let  $S$  be a Noetherian local ring with residue field  $k$ , and let  $M$  be a finitely generated  $S$ -module.

*Definition 3.1*

We define the *defect*  $d_S(M)$  of  $M$  to be

$$\begin{aligned} d_S(M) &= \dim_k \operatorname{Tor}_S^0(M, k) - \dim_k \operatorname{Tor}_S^1(M, k) \\ &= \dim_k M/\mathfrak{m}_S M - \dim_k \operatorname{Tor}_S^1(M, k). \end{aligned}$$

Let

$$\cdots \rightarrow P_i \rightarrow \cdots \rightarrow P_1 \rightarrow P_0 \rightarrow M \rightarrow 0$$

be a (possibly infinite) resolution of  $M$  by finite free  $S$ -modules. Assume that the image of  $P_i$  in  $P_{i-1}$  is contained in  $\mathfrak{m}_S P_{i-1}$  for each  $i \geq 1$ . (Such resolutions always exist and are often called *minimal*.) Let  $r_i$  denote the rank of  $P_i$ . Tensoring the resolution over  $S$  with  $k$  we see that  $P_i/\mathfrak{m}_S P_i \cong \operatorname{Tor}_S^i(M, k)$  and hence that  $r_i = \dim_k \operatorname{Tor}_S^i(M, k)$ .

*Definition 3.2*

We say that  $M$  is *balanced* if  $d_S(M) \geq 0$ .

If  $M$  is balanced, then we see that it admits a presentation

$$S^d \rightarrow S^d \rightarrow M \rightarrow 0$$

with  $d = \dim_k M/\mathfrak{m}_S M$ .

### 3.2. Patching

We recall the abstract Taylor–Wiles style patching result from [16].

#### PROPOSITION 3.3

*Suppose that*

- (1)  $R$  is an object of  $\mathcal{C}_\mathcal{O}$  and that  $H$  is a finite  $R$ -module which is also finite over  $\mathcal{O}$ ,
- (2)  $q \geq 1$  is an integer, and for each integer  $N \geq 1$ ,  $S_N := \mathcal{O}[(\mathbf{Z}/p^N \mathbf{Z})^q]$ ,
- (3)  $R_\infty := \mathcal{O}[[x_1, \dots, x_{q-1}]]$ ,
- (4) for each  $N \geq 1$ ,  $\phi_N : R_\infty \twoheadrightarrow R$  is a surjection in  $\mathcal{C}_\mathcal{O}$  and  $H_N$  is an  $R_\infty \otimes_\mathcal{O} S_N$ -module,

*and suppose that for each  $N \geq 1$  the following conditions are satisfied:*

- (a) *the image of  $S_N$  in  $\text{End}_\mathcal{O}(H_N)$  is contained in the image of  $R_\infty$ , and moreover, the image of the augmentation ideal of  $S_N$  in  $\text{End}_\mathcal{O}(H_N)$  is contained in the image of  $\ker(\phi_N)$ ;*
- (b) *there is an isomorphism  $\psi_N : (H_N)_{\Delta_N} \xrightarrow{\sim} H$  of  $R_\infty$ -modules, where  $R_\infty$  acts on  $H$  via  $\phi_N$  and  $\Delta_N = (\mathbf{Z}/p^N \mathbf{Z})^q$ ;*
- (c)  $H_N$  is finite and balanced over  $S_N$  (see Definition 3.2).

*Then  $H$  is a free  $R$ -module.*

*Proof*

This is [16, Proposition 2.3]. □

## 4. Deformations of Galois representations

Let

$$\bar{\rho} : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(k)$$

be a continuous, odd, absolutely irreducible Galois representation with similitude character of the form  $\nu(\bar{\rho}) = \bar{\epsilon}^{-(a-1)}$ , where  $a \geq 2$ . Let us suppose that there exist  $\alpha$  and  $\beta$  in  $k$  such that

$$\bar{\rho} |_{G_p} \sim \begin{pmatrix} \lambda(\alpha) & 0 & * & * \\ 0 & \lambda(\beta) & * & * \\ 0 & 0 & \nu(\bar{\rho}) \cdot \lambda(\beta^{-1}) & 0 \\ 0 & 0 & 0 & \nu(\bar{\rho}) \cdot \lambda(\alpha^{-1}) \end{pmatrix},$$

and moreover  $(\alpha^2 - 1)(\beta^2 - 1)(\alpha^2\beta^2 - 1)(\alpha - \beta) \neq 0$ . Let  $S(\bar{r})$  denote the set of primes of  $\mathbf{Q}$  away from  $p$  at which  $\bar{r}$  is ramified.

The group  $\mathrm{GSp}_4$  admits an 11-dimensional adjoint representation on its Lie algebra  $\mathfrak{g}$ . Let  $\mathrm{ad}(\bar{r})$  denote the composition of  $\bar{r}$  with this representation. For  $p > 2$ , the representation  $\mathrm{ad}(\bar{r})$  admits a decomposition  $\mathrm{ad}(\bar{r}) = \mathrm{ad}^0(\bar{r}) \oplus \nu$ , where  $\nu$  is the similitude character of  $\bar{r}$ .

We make the following further assumptions on  $\bar{r}$ .

*Assumption 4.1 (Big image)*

The restriction of  $\bar{r}$  to  $G_{\mathbf{Q}(\zeta_p)}$  satisfies the following conditions (cf. [56, Section 5.7]):

- H1: The field  $\mathbf{Q}(\mathrm{ad}^0(\bar{r}))$  does not contain  $\zeta_p$ .
- H2: For any  $m$ , there exists an element  $\sigma \in G_{\mathbf{Q}(\zeta_{p^m})}$  such that  $\bar{r}(\sigma)$  has four distinct eigenvalues and such that the action of  $\sigma$  on each irreducible representation of  $\mathrm{ad}^0(\bar{r})$  over  $G_{\mathbf{Q}(\zeta_{p^m})}$  contains 1 as an eigenvalue.
- H3: Neither the image  $\Gamma$  of  $\mathrm{ad}^0(\bar{r})$  nor the image of  $\mathrm{ad}^0(\bar{r})(1)$  admits a quotient of degree  $p$ .

If this assumption holds, we say that  $\bar{r}$  has *big image*, although condition (H1) depends on more than the group-theoretic image of  $\bar{r}$  or even  $\bar{r}|_{G_{\mathbf{Q}(\zeta_p)}}$ .

*Assumption 4.2 (Neatness)*

There exists a  $\sigma \in G_{\mathbf{Q}}$  with  $\epsilon(\sigma) = q \not\equiv 1 \pmod{p}$  such that the ratio of any two eigenvalues of  $\bar{r}(\sigma)$  is not equal to  $q \pmod{p}$ .

This condition is imposed to avoid dealing with stacks. If  $p \geq 5$ , then any surjective representation  $\bar{r}: G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathbf{F}_p)$  whose similitude character is a power of  $\bar{\epsilon}$  will be neat. By assumption, the image contains an element  $\bar{r}(\sigma)$  which is scalar with eigenvalue  $\lambda \neq \pm 1$ . If  $q = \epsilon(\sigma) \equiv 1 \pmod{p}$ , then the similitude character would also equal 1. But the similitude character of the scalar matrix  $\lambda$  is  $\lambda^2 \not\equiv 1 \pmod{p}$ .

*Assumption 4.3 (Ramification)*

If  $x \in S(\bar{r})$ , then  $\bar{r}|_{G_x}$  is one of the following types:

- (1) **U3**:  $\bar{r}|_{I_x}$  has unipotent image, and  $\bar{r}|_{I_x}$  is conjugate to the group generated by  $\exp(N_3)$ , where

$$N_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

- (2) **U2:**  $\bar{\tau} | I_x$  has unipotent image, and  $\bar{\tau} | I_x$  is conjugate to the group generated by  $\exp(N_2)$ , where

$$N_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

- (3) **U1:**  $\bar{\tau} | I_x$  has unipotent image, and  $\bar{\tau} | I_x$  is conjugate to the group generated by  $\exp(N_1)$ , where

$$N_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

- (4) **P:**  $\bar{\tau} | G_x$  is a direct sum of characters, and  $\bar{\tau} | I_x$  has the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \chi_x & 0 \\ 0 & 0 & 0 & \chi_x \end{pmatrix}$$

for some nontrivial character  $\chi_x$  of  $I_x$ . Both the plane of invariants under  $I_x$  and the plane on which  $I_x$  acts by  $\chi_x$  are isotropic. Moreover,  $x - 1$  is prime to  $p$ .

- (5) **H:**  $\bar{\tau} | I_x$  is absolutely irreducible, and  $x^4 - 1$  is prime to  $p$ .

*Remark 4.4*

Since we are assuming that the similitude character of  $\bar{\tau}$  is a power of the cyclotomic character, it turns out that  $\bar{\tau} | I_x$  can never be of type **P**. We expect that our arguments can also be adapted to deal with representations  $\bar{\tau}$  with more general (odd) similitude characters, but we made this assumption to simplify some of the arguments involving  $q$ -expansions (in particular, to avoid various Nebentypus characters).

Note that nontrivial unipotent representations are not direct sums, so a prime  $x \in S(\bar{\tau})$  is either of type **U**, **P**, or **H**, but never simultaneously any two of these types. Moreover,  $x$  is of type **U2** or **U3** if and only if  $\bar{\tau}(I_x)$  is generated by an element  $\exp(N)$ , where  $N$  is nilpotent of rank 2 or 3, respectively.

Let  $Q$  denote a finite set of primes of  $\mathbf{Q}$  disjoint from  $S(\bar{\tau}) \cup \{p\}$ . We assume that for each  $x \in Q$  the following hold:

- $x \equiv 1 \pmod{p}$ ,

- $\bar{\tau} | G_x$  is a direct sum of four pairwise distinct characters; label these characters as  $\lambda(\alpha_x)$ ,  $\lambda(\beta_x)$ ,  $\lambda(\gamma_x)$ , and  $\lambda(\delta_x)$  such that the planes  $\lambda(\alpha_x) \oplus \lambda(\beta_x)$  and  $\lambda(\gamma_x) \oplus \lambda(\delta_x)$  are isotropic and  $\alpha_x \delta_x = \beta_x \gamma_x = \nu(\bar{\tau})(\text{Frob}_x)$ .

(By abuse of notation, we sometimes use  $Q$  to denote the product of primes in  $\mathcal{C}_\theta$ .) For objects  $R$  in  $\mathcal{C}_\theta$ , a *deformation* of  $\bar{\tau}$  to  $R$  is a  $\ker(\text{GSp}_4(R) \rightarrow \text{GSp}_4(k))$ -conjugacy class of continuous lifts  $r : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(R)$  of  $\bar{\tau}$ . We will often refer to the deformation containing a lift  $r$  simply by  $r$ .

*Remark 4.5*

When deforming Galois representations over  $\mathbf{Q}$ , we could work either with a fixed or varying similitude character—both give rise to deformation problems with  $l_0 = 1$ . We make the (somewhat arbitrary) choice to work with deformations with fixed similitude character in this paper, because it is the “correct” approach for general totally real fields—for totally real fields other than  $\mathbf{Q}$ , the invariant  $l_0$  increases (by  $[F : \mathbf{Q}] - 1$ ) when deforming the similitude character.

*Definition 4.6*

We say that a deformation  $r : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(R)$  of  $\bar{\tau}$  is *minimal outside  $Q$*  if it satisfies the following properties:

- (1) The similitude character  $\nu(r)$  is equal to  $\epsilon^{-(a-1)}$ .
- (2) If  $x \notin Q \cup S(\bar{\tau}) \cup \{p\}$  is a prime of  $\mathbf{Q}$ , then  $r | G_x$  is unramified.
- (3) If  $x \in S(\bar{\tau})$  is of type **U1**, **U2**, or **U3**, then  $r | I_x$  has unipotent image and its image is topologically generated by an element  $\exp(N)$  where  $N$  is nilpotent of rank 1, 2, or 3, respectively.
- (4) If  $x \in S(\bar{\tau})$  is of type **P**, then  $r(I_x) \xrightarrow{\sim} \bar{\tau}(I_x)$ .
- (5) If  $x \in Q$ , then  $r | G_x \cong V_1 \oplus V_2$  where each  $V_i$  is an isotropic plane in  $R^4$  and  $V_1$  lifts  $\lambda(\alpha_x) \oplus \lambda(\beta_x)$  while  $V_2$  lifts  $\lambda(\gamma_x) \oplus \lambda(\delta_x)$ . Moreover,  $I_x$  acts by scalars (via some character) on  $V_1$  and by scalars via the inverse of this character on  $V_2$ .
- (6) The representation  $r$  has the following shape at  $p$ :

$$r | G_p \sim \begin{pmatrix} \chi_\alpha \psi^{-1} & 0 & * & * \\ 0 & \chi_\beta \psi^{-1} & * & * \\ 0 & 0 & \epsilon^{-(a-1)} \chi_\beta^{-1} \psi & 0 \\ 0 & 0 & 0 & \epsilon^{-(a-1)} \chi_\alpha^{-1} \psi \end{pmatrix},$$

where  $\chi_\alpha$  and  $\chi_\beta$  are unramified characters lifting  $\lambda(\alpha)$  and  $\lambda(\beta)$ , respectively, and  $\psi$  is an unramified character which is trivial modulo the maximal ideal.

If  $Q$  is empty, then we will refer to such deformations simply as being *minimal*. If  $r$  satisfies conditions (2)–(4), then we say that  $r$  is *weakly minimal* outside  $Q$ .

*Remark 4.7*

The local condition at  $p$  is equivalent to asking that  $r$  is ordinary (of fixed weight). When  $a = 2$  it is also equivalent to being finite flat. This is because, for unramified characters  $\psi_1$  and  $\psi_2$ , the group  $\text{Ext}^1(\psi_1, \psi_2)$  in this category is trivial, and the group  $\text{Ext}^1(\epsilon\psi_1, \psi_2)$  is the same whether it is computed in the category of finite flat group schemes or as  $G_p$ -modules, as long as  $\psi_1\psi_2^{-1} \not\equiv 1 \pmod{p}$ . The latter condition follows (for all the relevant extensions) from the assumption  $(\alpha\beta - 1)(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta) \neq 0$ .

The functor that associates to each object  $R$  of  $\mathcal{C}_\emptyset$  the set of deformations of  $\bar{r}$  to  $R$  which are minimal outside  $Q$  is represented by a complete Noetherian local  $\mathcal{O}$ -algebra  $R_Q$ . This follows from the proof of [21, Theorem 2.41]. If  $Q = \emptyset$ , we will sometimes denote  $R_Q$  by  $R^{\min}$ .

Let  $H_Q^1(\mathbf{Q}, \text{ad}^0(\bar{r}))$  denote the Selmer group defined as the kernel of the map

$$H^1(\mathbf{Q}, \text{ad}^0(\bar{r})) \longrightarrow \bigoplus_x H^1(\mathbf{Q}_x, \text{ad}^0(\bar{r}))/L_{Q,x},$$

where  $x$  runs over all primes of  $\mathbf{Q}$  and the following conditions are satisfied:

- If  $x \notin Q \cup p$ , then  $L_{Q,x} = H^1(G_x/I_x, (\text{ad}^0(\bar{r}))^{I_x})$ .
- If  $x \in Q$ , then  $H^1(G_x, \text{ad}^0(\bar{r}))$  is isomorphic to the subspace of

$$\begin{aligned} & H^1(G_x, \text{ad}\lambda(\alpha_x)) \oplus H^1(G_x, \text{ad}\lambda(\beta_x)) \oplus H^1(G_x, \text{ad}\lambda(\beta_x)^{-1}) \\ & \oplus H^1(G_x, \text{ad}\lambda(\alpha_x)^{-1}) \end{aligned}$$

consisting of elements  $(c_1, c_2, d_2, d_1)$  with  $c_1 + d_1 = c_2 + d_2$ . (Note that each summand is a copy of  $\text{Hom}_{\text{cts}}(G_x, k)$ .) We let  $L_{Q,x}$  denote the subspace corresponding to elements  $(c_1, c_2, d_2, d_1)$  with  $c_1 - c_2$  and  $d_1 - d_2$  and  $c_1 + d_1$  (equivalently,  $c_2 + d_2$ ) unramified.

- If  $x = p$ , then we define  $L_{Q,p} = L_p$  as follows: let  $\mathfrak{u} \subset \mathfrak{b}^0$  be the subspace of matrices whose nonzero entries appear in the upper right  $2 \times 2$  block. We define  $L'_p = \ker(H^1(G_p, \mathfrak{b}^0) \rightarrow H^1(I_p, \mathfrak{b}^0/\mathfrak{u}))$  and  $L_p = L_{Q,p} = \text{Im}(L'_p \rightarrow H^1(G_p, \mathfrak{g}^0))$ .

Let  $H_Q^1(\mathbf{Q}, \text{ad}^0(\bar{r}(1)))$  denote the corresponding dual Selmer group.

## LEMMA 4.8

We have  $\dim_k L_p - \dim_k H^0(G_p, \text{ad}^0(\bar{r})) = 3$ .

*Proof*

The subspace  $L'_p \subset H^1(G_p, \mathfrak{b}^0)$  is precisely a set of elements mapping to the subspace  $H^1(G_p/I_p, (\mathfrak{b}^0/\mathfrak{u})^{I_p}) \subset H^1(G_p, \mathfrak{b}^0/\mathfrak{u})$ . We have  $\mathfrak{b}^0/\mathfrak{u} \cong 1 \oplus 1 \oplus \lambda(\beta)\lambda(\alpha)^{-1}$



as a  $k[G_p]$ -module and hence  $H^1(G_p/I_p, (\mathfrak{b}^0/\mathfrak{u})^{I_p})$  is 2-dimensional since  $\alpha \neq \beta$ . The condition  $(\alpha^2 - 1)(\beta^2 - 1)(\alpha^2\beta^2 - 1)(\alpha - \beta) \neq 0$  implies that  $h^2(G_p, \mathfrak{u}) = 0$  and hence  $H^1(G_p, \mathfrak{b}^0) \twoheadrightarrow H^1(G_p, \mathfrak{b}^0/\mathfrak{u})$ . It follows that  $\dim_k L'_p = 2 + h^1(G_p, \mathfrak{b}^0) - h^1(G_p, \mathfrak{b}^0/\mathfrak{u})$ . Thus,

$$\dim_k L'_p - h^0(G_p, \mathfrak{b}^0) = 2 + h^1(G_p, \mathfrak{u}) - h^0(G_p, \mathfrak{b}^0/\mathfrak{u}) - h^0(G_p, \mathfrak{u}).$$

We have  $h^0(G_p, \mathfrak{u}) = 0$  and  $h^0(G_p, \mathfrak{b}^0/\mathfrak{u}) = 2$ . The Euler characteristic formula implies that  $h^1(G_p, \mathfrak{u}) = 3$ . Thus,

$$\dim_k L'_p - h^0(G_p, \mathfrak{b}^0) = 2 + 3 - 2 - 0 = 3.$$

Finally, the condition on  $\alpha$  and  $\beta$  implies that  $h^0(G_p, \mathfrak{g}^0/\mathfrak{b}^0) = 0$ . It follows that  $h^0(G_p, \mathfrak{b}^0) = h^0(G_p, \mathfrak{g}^0)$  and  $L'_p \xrightarrow{\sim} L_p$ . This concludes the proof.  $\square$

PROPOSITION 4.9

*The reduced tangent space  $\text{Hom}(R_Q/\mathfrak{m}_\emptyset, k[\epsilon]/\epsilon^2)$  of  $R_Q$  has dimension*

$$\dim_k H_Q^1(\mathbf{Q}, \text{ad}^0(\bar{\mathcal{F}}(1))) - 1 + \#Q.$$

*Proof*

The argument is very similar to that of [21, Corollary 2.43]. The reduced tangent space has dimension  $\dim_k H_Q^1(\mathbf{Q}, \text{ad}^0(\bar{\mathcal{F}}))$ . By [21, Theorem 2.18], this is equal to

$$\begin{aligned} & \dim_k H_Q^1(\mathbf{Q}, \text{ad}^0(\bar{\mathcal{F}}(1))) + \dim_k H^0(\mathbf{Q}, \text{ad}^0(\bar{\mathcal{F}})) - \dim_k H^0(\mathbf{Q}, \text{ad}^0(\bar{\mathcal{F}}(1))) \\ & + \sum_x (\dim_k L_{Q,x} - \dim_k H^0(\mathbf{Q}_x, \text{ad}^0(\bar{\mathcal{F}}))) - \dim_k H^0(G_\infty, \text{ad}^0(\bar{\mathcal{F}})), \end{aligned}$$

where  $x$  runs over all finite places of  $\mathbf{Q}$ . The second term is equal to zero and the third term vanishes (by the absolute irreducibility of  $\bar{\mathcal{F}}$  and the fact that  $\bar{\mathcal{F}} \not\cong \bar{\mathcal{F}} \otimes \epsilon$ ). Now, we have

- $\dim_k L_{Q,x} - \dim_k H^0(\mathbf{Q}_x, \text{ad}^0(\bar{\mathcal{F}})) = 0$  for  $x \notin Q \cup \{p\}$ ,
- $\dim_k L_{Q,x} - \dim_k H^0(\mathbf{Q}_x, \text{ad}^0(\bar{\mathcal{F}})) = 3$  for  $x = p$ ,
- $\dim_k L_{Q,x} - \dim_k H^0(\mathbf{Q}_x, \text{ad}^0(\bar{\mathcal{F}})) = 1$  for  $x \in Q$  (by [31, Proposition 10.4.1]), and
- $\dim_k H^0(G_\infty, \text{ad}^0(\bar{\mathcal{F}})) = 4$ .

This concludes the proof.  $\square$

The next result (on the existence of Taylor–Wiles primes) follows from the previous proposition and the proof of [56, Proposition 5.6].

## PROPOSITION 4.10

Let  $q = \dim_k H_{\mathcal{O}}^1(\mathbf{Q}, \text{ad}^0(\bar{r}(1)))$ , and recall that we are supposing  $\bar{r}$  satisfies Assumption 4.1. Then  $q \geq 1$  and for any integer  $N \geq 1$  we can find a set  $Q_N$  of primes of  $\mathbf{Q}$  such that

- (1)  $\#Q_N = q$ ,
- (2)  $x \equiv 1 \pmod{p^N}$  for each  $x \in Q_N$ ,
- (3) for each  $x \in Q_N$ ,  $\bar{r}$  is unramified at  $x$  and  $\bar{r}(\text{Frob}_x)$  has four pairwise distinct eigenvalues, and
- (4)  $H_{Q_N}^1(\mathbf{Q}, \text{ad}(\bar{r}(1))) = (0)$ .

In particular, the reduced tangent space of  $R_{Q_N}$  has dimension  $q - 1$  and  $R_{Q_N}$  is a quotient of a power series ring over  $\mathcal{O}$  in  $q - 1$  variables.

## Example 4.11 (Examples of representations with big image)

Suppose that  $p \geq 5$ .

- (1) Let  $K/\mathbf{Q}$  be an imaginary quadratic field not contained in  $\mathbf{Q}(\zeta_p)$ . Let

$$\bar{\rho}: G_K \rightarrow \text{GL}_2(\mathbf{F}_p)$$

be a representation with determinant  $\epsilon^{1-k}$  for some integer  $k$  such that the images of  $\bar{\rho}$  and  $\bar{\rho}^c$  for any complex conjugation  $c \in \text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$  both contain  $\text{SL}_2(\mathbf{F}_p)$  and have totally disjoint fixed fields over  $K(\zeta_p)$ . Then the representation

$$\bar{r} = \text{Ind}_K^{\mathbf{Q}} \bar{\rho}$$

preserves a symplectic form and has big image.

- (2) Suppose that the image of  $\bar{r}$  is  $\text{GSp}_4(\mathbf{F}_p)$ . Then  $\bar{r}$  has big image.

*Proof*

The second claim follows immediately for  $p \geq 5$  by [56, Proposition 5.8]. For the first claim, it is an easy consequence of the fact that  $\text{SL}_2(\mathbf{F}_p)$  is perfect for  $p \geq 5$  that  $H3$  holds, and similarly, assuming that  $K \not\subset \mathbf{Q}(\zeta_p)$ , that  $H1$  holds. Hence it suffices to find an element in the image with distinct eigenvalues and with 1 as an eigenvalue for every irreducible constituent of  $\text{ad}^0(\bar{r})$ . We first compute the representation  $\text{ad}^0(\bar{r})$ . Note that the dual of  $\bar{\rho}$  and  $\bar{\rho}^c$  can be identified with  $\bar{\rho} \times \epsilon^{k-1}$  and  $\bar{\rho}^c \otimes \epsilon^{k-1}$ , respectively. Over  $K$ , we have an identification

$$\text{ad}^0(\bar{r})|_{G_K} = (\bar{\rho} \otimes \bar{\rho}^c) \otimes \epsilon^{k-1} \oplus \text{ad}^0(\bar{\rho}) \oplus \text{ad}^0(\bar{\rho}^c),$$

and over  $\mathbf{Q}$ , we have an identification

$$\text{ad}^0(\bar{r}) = \text{As}(\bar{\rho}) \otimes \epsilon^{k-1} \oplus \text{Ind}_K^{\mathbf{Q}} \text{ad}^0(\bar{\rho}),$$

where  $\text{As}$  is the Asai representation. Over  $\mathbf{Q}(\zeta_{p^m})$  for any  $m$ , the character  $\epsilon^{k-1}$  is trivial, and hence the image of  $\bar{r}|_{G_{\mathbf{Q}(\zeta_{p^m})}}$  under our assumptions is the group  $\text{SL}_2(\mathbf{F}_p)^2 \rtimes \mathbf{Z}/2\mathbf{Z}$ . Since 1 and  $-1$  are always eigenvalues of any element acting on  $\text{Ind}_K^{\mathbf{Q}} \text{ad}^0(\bar{\rho})$ , it suffices to find an element  $\sigma \in \text{SL}_2(\mathbf{F}_p)^2 \rtimes \mathbf{Z}/2\mathbf{Z}$  which has distinct eigenvalues under  $\bar{r}$  and has an eigenvalue 1 in  $\text{As}(\bar{\rho})$ . To be more precise, since we have not been careful about distinguishing the Asai representation from its quadratic twist, we will find an element with eigenvalues both 1 and  $-1$ . One can explicitly realize the Asai representation as follows. Let  $V$  be the standard representation of  $\text{SL}_2(\mathbf{F}_p)$  over  $\mathbf{F}_p$ , and let  $V \otimes V$  be the representation of the exterior product  $\text{SL}_2(\mathbf{F}_p) \times \text{SL}_2(\mathbf{F}_p)$ . The element  $(g, h)$  acts on  $v \otimes w$  via  $(g, h)(v \otimes w) = (gv \otimes hw)$ . The Asai representation is determined uniquely by the action of a fixed lift of complex conjugation  $c \in \text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ , which acts on  $V \otimes V$  by the formula  $c(v \otimes w) = w \otimes v$ .

Consider the elements  $g, h \in \text{SL}_2(\mathbf{F}_p)$  such that, with respect to some chosen basis  $V = \{u, v\}$ ,

$$g = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix}, \quad h = \begin{pmatrix} y & 0 \\ 0 & y^{-1} \end{pmatrix}.$$

Then  $c \cdot (g, h)$  acts on  $\bar{r}$  via the matrix

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x & 0 & 0 & 0 \\ 0 & x^{-1} & 0 & 0 \\ 0 & 0 & y & 0 \\ 0 & 0 & 0 & y^{-1} \end{pmatrix}$$

with eigenvalues  $\pm(xy)^{1/2}$  and  $\pm(xy)^{-1/2}$ . On the other hand, the action of this element via the Asai representation (and basis  $u \otimes u, v \otimes v, u \otimes v, v \otimes u$ ) is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} xy & 0 & 0 & 0 \\ 0 & (xy)^{-1} & 0 & 0 \\ 0 & 0 & (x/y) & 0 \\ 0 & 0 & 0 & (x/y)^{-1} \end{pmatrix}$$

with eigenvalues  $xy, (xy)^{-1}$ , and  $\pm 1$ . The four eigenvalues are distinct as long as  $\pm(xy)^{1/2} \neq \pm(xy)^{-1/2}$ , or equivalently, if  $(xy)^2 \neq 1$ . One can now choose  $x = 2$  and  $y = 1$  in  $\mathbf{F}_p^\times$ .  $\square$

*Remark 4.12*

Suppose that  $K$  is an imaginary quadratic field, and suppose that  $E/K$  is an elliptic curve which neither has CM nor is isogenous (over  $\bar{K}$ ) to its Galois conjugate  $E^c/K$ . We claim that Example 4.11 applies to the mod  $p$  representations  $\bar{\rho}: G_K \rightarrow \text{GL}_2(\mathbf{F}_p)$  associated to the dual of  $E[p]$  for sufficiently large  $p$ . The representations  $\bar{r}$  in this

case are the duals of the representations  $A[p]$  associated to the abelian surface  $A = \text{Res}_K^{\mathbb{Q}}(E)$ . By [61], the Galois representations  $\bar{\rho}_p, \bar{\rho}_p^c : G_K \rightarrow \text{GL}_2(\mathbf{F}_p)$  associated to the duals of  $E[p]$  and  $E^c[p]$  have images  $\text{GL}_2(\mathbf{F}_p)$  and determinants  $\epsilon^{1-2}$  for all sufficiently large  $p \geq 5$ . Let  $F/K$  and  $F^c$  denote the corresponding extensions, so  $\text{Gal}(F/K)$  and  $\text{Gal}(F^c/K)$  are both isomorphic to  $\text{GL}_2(\mathbf{F}_p)$ , and  $\text{Gal}(F/K(\zeta_p))$  and  $\text{Gal}(F^c/K(\zeta_p))$  are both isomorphic to  $\text{SL}_2(\mathbf{F}_p)$ . By the simplicity of  $\text{PSL}_2(\mathbf{F}_p)$  for  $p \geq 5$ , the only nontrivial quotients of  $\text{SL}_2(\mathbf{F}_p)$  are  $\text{PSL}_2(\mathbf{F}_p)$  and  $\text{SL}_2(\mathbf{F}_p)$ . This implies that if  $H := F \cap F^c \supseteq K(\zeta_p)$  is strictly larger than  $K(\zeta_p)$ , then either  $\text{Gal}(H/K) = \text{GL}_2(\mathbf{F}_p)$  or  $\text{Gal}(H/K) = \text{GL}_2(\mathbf{F}_p)/\pm I$ . In either case, the projective representations associated to  $\bar{\rho}_p$  and  $\bar{\rho}_p^c$  both factor through  $\text{Gal}(H/K)$ . Since all automorphisms of  $\text{PGL}_2(\mathbf{F}_p)$  are inner, this implies that projective representations of  $\bar{\rho}_p$  and  $\bar{\rho}_p^c$  are isomorphic, and hence  $\bar{\rho}_p \simeq \bar{\rho}_p^c \otimes \chi_p$  for some character  $\chi_p$  which (by comparing determinants) is at most quadratic. Assume that  $p$  is sufficiently large so that  $E$  has good reduction at all primes above  $p$  and, moreover, that  $p$  is unramified in  $K$ . Then  $\bar{\rho}_p$  and  $\bar{\rho}_p^c$  are both finite flat at  $v \mid p$ , which forces  $\chi_p$  to be unramified at all primes above  $p$ . But this implies that  $\chi_p$  is unramified outside primes dividing the conductor  $N$  and  $N^c$  of  $E$  and  $E^c$ , respectively. There are only finitely many such quadratic characters by class field theory. Hence, if there are infinitely many primes  $p$  for which the assumptions of Example 4.11 do not occur, then there exists a *fixed* character  $\chi$  with  $\chi^2 = 1$  and isomorphisms  $\bar{\rho}_p \simeq \bar{\rho}_p^c \otimes \chi$  for infinitely many  $p$ . Such an isomorphism (for a single  $p$ ) implies that  $a_v = \chi(v)a_{v^c} \pmod p$  for all pairs of conjugate primes  $v$  and  $v^c$  of good reduction for  $E$ , and hence, given infinitely many such  $p$ , one deduces the equality  $a_v = \chi(v)a_{v^c}$ . If  $L/K$  is the (at most) quadratic extension in which  $\chi$  splits, this implies (by Chebotarev) that the Tate modules (for any fixed prime) of  $E$  and  $E^c$  are isomorphic, and hence (by Faltings [25]) that  $E$  and  $E^c$  are isogenous over  $L$ .

## 5. Siegel 3-folds

### 5.1. Level structure

Recall that there are two conjugacy classes of maximal parabolic subgroups of  $\text{GSp}(4)$  represented by the *Siegel parabolic*  $P$  which is block upper triangular with Levi

$$M = M_P := \left\{ \begin{pmatrix} A & & & \\ & \lambda B & & \\ & & S^t A^{-1} S & \\ & & & 1 \end{pmatrix} : \lambda \in \text{GL}_1, A \in \text{GL}_2, B = S^t A^{-1} S, S = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \right\},$$

and the *Klingen parabolic*  $\Pi$  which is block upper triangular with Levi

$$M_{\Pi} := \left\{ \begin{pmatrix} \lambda & & & \\ & A & & \\ & & \lambda^{-1} \det(A) & \\ & & & \end{pmatrix} : \lambda \in \text{GL}_1, A \in \text{GL}_2 \right\}.$$

These both contain the Borel subgroup  $B$ . For each prime  $x$ , these give rise to parahoric subgroups  $P(x)$ ,  $\Pi(x)$ , and  $I(x)$  of  $\mathrm{GSp}_4(\mathbf{Z}_x)$ , namely, the inverse image of the corresponding parabolic subgroups over  $\mathbf{F}_x$ . (The group  $I(x)$  is called the *Iwahori subgroup*.) The Klingen parahoric subgroup contains a normal subgroup  $\Pi(x)^+$  with  $\Pi(x)/\Pi(x)^+ \simeq (\mathbf{Z}/x\mathbf{Z})^\times$  (via projection onto  $\lambda \bmod x$ ). For each prime  $x$ , we also have the *paramodular group*  $K(x)$ , which is the stabilizer in  $\mathrm{GSp}_4(\mathbf{Q}_x)$  of  $\mathbf{Z}_x \oplus \mathbf{Z}_x \oplus \mathbf{Z}_x \oplus x\mathbf{Z}_x$ , and is the intersection

$$\left( \begin{array}{cccc} * & * & * & */x \\ * & * & * & */x \\ * & * & * & */x \\ x* & x* & x* & x \end{array} \right) \cap \mathrm{GSp}_4(\mathbf{Q}_x)$$

for values  $* \in \mathbf{Z}_x$ .

### 5.2. Cohomology of Siegel 3-folds

Let  $S$  and  $Q$  be finite sets of primes of  $\mathbf{Q}$  which are disjoint from each other and do not contain  $p$ . By a slight abuse of notation, we will sometimes denote the product of the primes in  $Q$  by the same symbol  $Q$ . For each  $x \in S$ , let  $K_x \subset \mathrm{GSp}_4(\mathbf{Z}_x)$  equal one of  $S(x)$ ,  $\Pi(x)$ ,  $K(x)$ ,  $\Pi(x)^+$ ,  $I(x)$  or the full congruence subgroup of level  $x$ . For  $x \notin S$ , we let  $K_x = \mathrm{GSp}_4(\mathbf{Z}_x)$  and define  $K := \prod_x K_x \subset \mathrm{GSp}_4(\mathbb{A}^\infty)$ . For  $x \in Q$ , we let  $K_{x,0} = \Pi(x)$  and  $K_{x,1} = \Pi^+(x)$ . Let  $K_i(Q) = \prod_{x \notin Q} K_x \times \prod_{x \in Q} K_{x,i}$  for  $i = 0, 1$ .

We assume that the subgroup  $K$  is neat. (This will be the case if  $S$  contains a prime  $x \geq 3$  where  $K_x$  is the full congruence subgroup of level  $x$ .) We let  $Y_K \rightarrow \mathrm{Spec}(\mathcal{O})$  (resp.,  $Y_{K_i(Q)} \rightarrow \mathrm{Spec}(\mathcal{O})$ ) denote the Siegel moduli space of level  $K$  (resp.,  $K_i(Q)$ ). This scheme classifies principally polarized abelian varieties together with a  $K$ -level structure (resp.,  $K_i(Q)$ -level structure). (See [57, Section 4.1].) In each case we denote the universal abelian variety by  $\mathcal{A}$ .

If  $Y$  denotes one of the above spaces, then we can choose a toroidal compactification  $X \rightarrow \mathrm{Spec}(\mathcal{O})$  of  $Y$ . The abelian scheme  $\mathcal{A}$  then extends to a semi-abelian scheme  $\pi : \mathcal{A} \rightarrow X$ , and the sheaf  $\mathcal{E} := \pi_* \Omega_{\mathcal{A}/X}^1$  is a locally free  $\mathcal{O}_X$ -module of rank 2. For integers  $a \geq b$ , we let  $\omega(a, b) := \mathrm{Sym}^{a-b} \mathcal{E} \otimes \det^b \mathcal{E}$ . We also denote  $\det \mathcal{E}$  by  $\omega$ , so, for example,  $\omega(a, a) = \omega^a$  is a line bundle. If  $M$  is an  $\mathcal{O}$ -module, then we will let  $\omega(a, b)_M$  denote the sheaf  $\omega(a, b) \otimes_{\mathcal{O}} M$ . The coherent cohomology groups  $H^i(X, \omega(a, b)_M)$  are independent of the choice of toroidal compactification  $X$  (see [51, Lemma 7.1.1.4] and the proof of [51, Lemma 7.1.1.5]). The Koecher principle states that there is an isomorphism

$$H^0(Y, \omega(a, b)_M) \simeq H^0(X, \omega(a, b)_M).$$

We may therefore pass freely between the open variety  $Y$  and the (any) smooth projective toroidal compactification  $X$  without comment when dealing with  $H^0$ .

We choose toroidal compactifications  $X_K$  and  $X_{K_0(Q)}$  so that the natural map  $Y_{K_0(Q)} \rightarrow Y_K$  extends to a map  $X_{K_0(Q)} \rightarrow X_K$ . As explained in [57, Section 4.1.2], the universal subgroup  $H \subset \mathcal{A}[Q]$  over  $Y_{K_0(Q)}$  extends to  $X_{K_0(Q)}$ . We then define the toroidal compactification  $X_{K_1(Q)} = \text{Isom}_{X_{K_0(Q)}}(\mathbf{Z}/Q, H)$ . The resulting map  $X_{K_1(Q)} \rightarrow X_{K_0(Q)}$  is then finite étale with Galois group  $\Delta_Q := (\mathbf{Z}/Q)^\times$ .

### 5.3. Vanishing results

Let  $X$  denote one of the toroidal compactifications defined in Section 5.2. We first record some consequences of a vanishing theorem of Lan and Suh.

#### THEOREM 5.1

(1) *Suppose that  $a \geq 3$  and  $2 \leq a - b \leq p - 2$ . Then*

$$H^i(X, \omega(a, b)(-\infty)_k) = 0$$

*for  $i > 2$ .*

(2) *Suppose that  $a + b \geq 6$  and  $2 \leq a - b \leq p - 2$ . Then*

$$H^i(X, \omega(a, b)(-\infty)_k) = 0$$

*for  $i > 1$ .*

(3) *Suppose that  $b \geq 4$  and  $0 \leq a - b \leq p - 4$ . Then*

$$H^i(X, \omega(a, b)(-\infty)_k) = 0$$

*for  $i > 0$ .*

#### *Proof*

This follows from [53, Corollary 7.24] after unwinding definitions. We take the group scheme  $G_1/R_1$  (in the notation of [53]) to be our  $G/\mathcal{O}$ . The groups  $M_1 \subset P_1 \subset G_1$  correspond to the Siegel Levi and parabolic:  $M \subset P \subset G$ . The set of dominant weights  $X_{G_1}^+$  (resp.,  $X_{M_1}^+$ ) is our  $X^*(T)_G^+$  (resp.,  $X^*(T)_M^+$ ) from Definition 2.1.

In this paragraph, we show that the subset  $X_{G_1}^{+, <reP} \subset X_{G_1}^+$  as defined in [52, Definition 6.3] corresponds to the set of those  $\mu = (a, b; c) \in X^*(T)_G^+$  such that  $a + b < p - 3$ . As an intermediate step, we first show that  $X_{G_1}^{+, <reP}$  corresponds to those  $\mu = (a, b; c) \in X^*(T)_G^+$  such that

- $\langle \mu + \rho, \pm \alpha_i^\vee \rangle \leq p$  for  $i = 1, \dots, 4$ ,
- $a + b + 3 < p$ .

To see this, we note the following: to lie in  $X_{G_1}^{+, <reP}$ , by definition, the element  $\mu$  must satisfy  $|\mu|_L + d < p$  and must also lie in  $X_{G_1}^{+, <wP}$ . The definition of  $|\mu|_L$  in

[52, Definition 3.2] boils down to  $|\mu|_L = a + b$  (the set  $\Upsilon$  in our case consists of the single embedding  $\mathbf{Z} \hookrightarrow \mathcal{O}$  and the norm  $|\mu| = a + b$  is defined near the beginning of Section 2.5). The dimension  $d$  is defined in [52, Definition 3.9] to be  $\dim_{\mathcal{O}}(X)$  which is 3 in our case. Next, the set  $X_{G_1}^{+, <wp}$  is defined in [52, Definition 3.2] to consist of those  $\mu \in X_{G_1}^{+, <p}$  for which  $|\mu|_L < p$ . Finally, the set  $X_{G_1}^{+, <p}$  is defined in [52, Definition 2.29] to consist of all dominant  $\mu \in X_{G_1}^+$  which satisfy the first condition above. This establishes the intermediate step. Now, if  $\mu \in X^*(T)_G^+$ , then the largest of the  $\langle \mu + \rho, \pm \alpha_i^\vee \rangle$  is  $\langle \mu + \rho, \alpha_3^\vee \rangle = a + b + 3$ . Thus, we see that  $\mu \in X_{G_1}^{+, <wp}$  if and only if  $a \geq b \geq 0$  and  $a + b < p - 3$ .

The set  $X_{M_1}^{+, <p}$ , by [52, Definition 2.29], is  $\{\mu \in X^*(T)_M^+ : \langle \mu + \rho, \alpha_1^\vee \rangle \leq p\} = \{(a, b; c) \in X^*(T)_M^+ : (a + 2) - (b + 1) \leq p\}$ . By Lemma 7.2, Definition 7.14 (which is vacuous in our case), and Proposition 7.15 of [52], a weight  $\mu = (a, b; c)$  lies in  $X_{M_1}^{+, <p}$  and is *positive parallel* if and only if  $a = b > 0$ .

If  $\mu = (a, b; c) \in X^*(T)_M^+$ , then a pair of vector bundles  $\mathcal{W}_\mu^*$ , for  $*$   $\in \{\text{can}, \text{sub}\}$  is defined in [53]. Indeed  $\mu$  determines an algebraic representation of  $M \cong \text{GL}_2 \times \text{GL}_1$  over  $\mathcal{O}$  with highest weight  $(a, b; c)$  (namely  $(\text{Sym}^{a-b} S_2 \otimes \det^b S_2) \otimes S_1^{\otimes c}$ , where  $S_i$  is the standard representation of  $\text{GL}_i$ ) and the corresponding bundles are then defined by [53, Definition 4.12]. We claim that

$$\mathcal{W}_\mu^{\text{can}} = \omega(a, b).$$

(We note that the parameter  $c$  does not change the underlying vector bundle, but does change the Hecke action on cohomology by a power of the similitude character.) Let  $\mu = (0, -1; 1)$ , let  $L$  denote the standard representation of  $G$ , and let  $L_0^\vee(1) \subset L$  be the subspace spanned by the first two standard basis vectors. Then  $L_0^\vee(1)$  is the standard representation of the  $\text{GL}_2$ -factor of  $M$  and is the representation of  $M$  corresponding to  $(1, 0; 0)$ . The representation  $L_0 = (L_0^\vee(1))^\vee(1)$  thus corresponds to  $\mu = (0, -1; 1)$ . By [52, Example 1.22], we have (in the notation of that paper)  $\mathcal{E}_{M_1}(L_0) = \text{Lie}_{\mathcal{A}/Y}$ . However,  $\mathcal{W}_\mu = \mathcal{E}_{M_1}(L_0)$  by definition, and we have  $\text{Lie}_{\mathcal{A}/Y} = \mathcal{E}^\vee = \omega(0, -1)$ . It follows that  $\mathcal{W}_{(0, -1; 1)}^{\text{can}} \cong \omega(0, -1)$ . We deduce that  $\omega(a, b) = (\text{Sym}^{a-b} \otimes \det^b)(\omega(1, 0)) = \mathcal{W}_{(a, b; -a-b)}$ , as required.

With these preliminaries out of the way, we now apply [53, Corollary 7.24]. We take  $\mu = (\alpha, \beta; \gamma) \in X_{G_1}^{+, <wp}$ . (The condition that  $\max(2, r_\tau) < p$  when  $\tau = \tau \circ c$  boils down to  $2 < p$  in our case.) We take  $\nu = (t, t; 0)$  a positive parallel weight. We therefore have  $t > 0$ ,  $\alpha \geq \beta \geq 0$  and  $\alpha + \beta < p - 3$ .

We now apply part 2 of [53, Corollary 7.24] successively with  $w \in W^{M_1}$  taken to equal each of the elements  $\tilde{w}_1, \tilde{w}_2, \tilde{w}_3$  from Section 2. Note that each  $\tilde{w}_i$  has length  $i$ . If we take  $w = \tilde{w}_1$ , then (ignoring the third component)

$$\tilde{w}_1 \cdot \mu - \nu = \tilde{w}_1(\alpha + 2, \beta + 1) - (2, 1) - (t, t) = (\alpha - t, -\beta - 2 - t).$$

Thus  $(\mathcal{W}_{\tilde{w}_1, \mu - \nu}^\vee)^{\text{sub}} = \omega(\beta + 2 + t, -\alpha + t)(-\infty)$ . Then [53, Corollary 7.24] implies that

$$H^i(X, \omega(\beta + 2 + t, -\alpha + t)(-\infty)_k) = 0$$

for each  $i > 2$ . Taking  $a = \beta + 2 + t$  and  $b = -\alpha + t$  gives the first part of our proposition.

Similarly, if  $w = \tilde{w}_2$ , then

$$\tilde{w}_2 \cdot \mu - \nu = \tilde{w}_2(\alpha + 2, \beta + 1) - (2, 1) - (t, t) = (\beta - 1 - t, -\alpha - 3 - t).$$

Hence

$$H^i(X, \omega(\alpha + 3 + t, 1 - \beta + t)(-\infty)_k) = 0$$

for  $i > 1$ . This gives the second part of the proposition.

Finally, we take  $w = \tilde{w}_3$ . Then

$$\tilde{w}_3 \cdot \mu - \nu = \tilde{w}_3(\alpha + 2, \beta + 1) - (2, 1) - (t, t) = (-\beta - 3 - t, -\alpha - 3 - t).$$

Hence

$$H^i(X, \omega(\alpha + 3 + t, \beta + 3 + t)(-\infty)_k) = 0$$

for  $i > 0$ . This gives the last part of the proposition.  $\square$

It is interesting to compare the above vanishing result in characteristic  $p$  with the following characteristic 0 vanishing results due (in part) to Blasius, Harris, and Ramakrishnan [9]. We have an identification

$$Y(\mathbf{C}) = G(\mathbf{Q}) \backslash (G(\mathbf{R})/K^h \times G(\mathbb{A}^\infty)/U),$$

where  $U \subset G(\mathbb{A}^\infty)$  is the open compact subgroup used to define  $Y$  and  $K^h$  is the compact-mod-center subgroup defined in Section 2.2. To any finite-dimensional  $\mathbf{C}$ -representation  $(\sigma, V_\sigma)$  of  $K_{\mathbf{C}}^h$ , there is an associated vector bundle  $\mathcal{V}_\sigma$  on  $Y(\mathbf{C})$  which is defined in [9, Definition 1.3.2]. This bundle has extensions  $\mathcal{V}_\sigma^{\text{sub}} \subset \mathcal{V}_\sigma^{\text{can}}$  to  $X(\mathbf{C})$ . In [9], the bundle  $\mathcal{V}_\sigma^{\text{can}}$  is denoted  $\tilde{\mathcal{V}}_\sigma$ . We have  $\mathcal{V}_\sigma^{\text{sub}} = \mathcal{V}_\sigma^{\text{can}}(-\infty)$ . For each  $i \geq 0$ , we define

$$\overline{H}^i(X(\mathbf{C}), \mathcal{V}_\sigma) := \text{Im}(H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{sub}}) \rightarrow H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{can}})).$$

Let  $\tilde{H}^i(\mathcal{V}_\sigma^{\text{sub}})$  and  $\tilde{H}^i(\mathcal{V}_\sigma^{\text{can}})$  denote the direct limit of  $H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{sub}})$  and  $H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{can}})$ , respectively, over all levels  $K$ . Let  $\overline{H}^i(X(\mathbf{C}), \mathcal{V}_\sigma)$  denote the corresponding



limit of  $\overline{H}^i(X(\mathbf{C}), \mathcal{V}_\sigma)$  (including both an overline and a tilde in the notation was too cumbersome; hopefully no confusion will result).

Let  $\mathcal{A}_{(2)}(G)$  denote the space of automorphic forms on  $G(\mathbf{Q}) \backslash G(\mathbb{A})$  which are square-integrable modulo the center  $Z_G(\mathbb{A})$ . Let  $\mathcal{A}_0(G) \subset \mathcal{A}_{(2)}(G)$  denote the space of cusp forms. For  $(\sigma, V_\sigma)$  a representation of  $K_{\mathbf{C}}^h$  as above and  $i \geq 0$ , we define

$$\begin{aligned}\mathcal{H}_{(2),\sigma}^i &= H^i(\mathrm{Lie} Q^-, K^h; \mathcal{A}_{(2)}(G) \otimes V_\sigma), \\ \mathcal{H}_{\mathrm{cusp},\sigma}^i &= H^i(\mathrm{Lie} Q^-, K^h; \mathcal{A}_0(G) \otimes V_\sigma).\end{aligned}$$

Then we have the following result of Harris.

**THEOREM 5.2**

*There are canonical maps, forming a commutative diagram*

$$\begin{array}{ccc}\mathcal{H}_{\mathrm{cusp},\sigma}^i & \longrightarrow & \mathcal{H}_{(2),\sigma}^i \\ \downarrow & & \downarrow \\ \widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{sub}}) & \longrightarrow & \widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})\end{array}$$

Moreover:

- (1) *The composition  $\mathcal{H}_{\mathrm{cusp},\sigma}^i \rightarrow \overline{H}^i(\mathcal{V}_\sigma)$  is injective for all  $i$ , and is an isomorphism for  $i = 0, 3$ .*
- (2) *The image of  $\mathcal{H}_{(2),\sigma}^i$  in  $\widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})$  contains  $\overline{H}^i(\mathcal{V}_\sigma)$ .*

*Proof*

This follows from [37, Theorem 2.7, Proposition 3.2.2]. □

For  $* \in \{\mathrm{cusp}, (2)\}$ , we then define  $\widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})_*$  to be the image of the space  $\mathcal{H}_{*,\sigma}^i$  in  $\widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})$ . Thus, we have

$$\mathcal{H}_{\mathrm{cusp},\sigma}^i \cong \widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})_{\mathrm{cusp}} \subset \overline{H}^i(\mathcal{V}_\sigma) \subset \widetilde{H}^i(\mathcal{V}_\sigma^{\mathrm{can}})_{(2)}.$$

For  $* \in \{\mathrm{cusp}, (2)\}$ , the space  $\mathcal{A}_*(G)$  is semisimple as a  $G(\mathbb{A})$ -representation and we decompose

$$\mathcal{A}_*(G) = \bigoplus_{\pi} m_*(\pi) \pi^\infty \otimes \pi_\infty.$$

We let  $\mathcal{A}_*(G)_{\mathrm{temp}}$  denote the subspace

$$\bigoplus m_*(\pi) \pi^\infty \otimes \pi_\infty,$$

where the sum is over all those  $\pi$  such that  $\pi_\infty$  is essentially tempered. We define  $\mathcal{H}_{*,\sigma,\text{temp}}^i \subset \mathcal{H}_{*,\sigma}^i$  by replacing  $\mathcal{A}_*(G)$  with  $\mathcal{A}_*(G)_{\text{temp}}$  in the definition of  $\mathcal{H}_{*,\sigma}^i$ . We then define

$$\widetilde{H}^i(\mathcal{V}_\sigma^{\text{can}})_{*,\text{temp}} \subset \widetilde{H}^i(\mathcal{V}_\sigma^{\text{can}})_*$$

to be the image of  $\mathcal{H}_{*,\sigma,\text{temp}}^i \rightarrow \widetilde{H}^i(\mathcal{V}_\sigma^{\text{can}})$ . We may also define analogous spaces

$$H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{can}})_{*,\text{temp}} \subset H^i(X(\mathbf{C}), \mathcal{V}_\sigma^{\text{can}})_*$$

by applying  $K$ -invariants to the constructions above, where  $K$  is the level of  $X(\mathbf{C})$ .

Suppose now that  $(\sigma, V_\sigma)$  is the irreducible representation of  $K_{\mathbf{C}}^h$  of highest weight  $\mu = (a, b; c) \in X^*(H_{\mathbf{C}})$ , with respect to the system of positive weights fixed in Section 2. We first of all observe that the bundle  $\mathcal{V}_\sigma$  does not depend on  $c$ . Indeed, let  $(\tau, V_\tau)$  be the irreducible representation of highest weight  $(a, b; c + 2)$ . Consider the  $G(\mathbf{Q})$ -equivariant bundles  $\mathcal{V}_\sigma^\vee = G(\mathbf{C}) \times_{Q^-} V_\sigma$  and  $\mathcal{V}_\tau^\vee = G(\mathbf{C}) \times_{Q^-} V_\tau$  on  $G(\mathbf{C})/Q^-$  defined in [9, Section 1.3]. (The superscripted  $\vee$ 's do not refer to dual bundles here.) Then by the definition of  $\mathcal{V}_\sigma$ , it suffices to show that  $\mathcal{V}_\sigma^\vee \xrightarrow{\sim} \mathcal{V}_\tau^\vee$  as  $G(\mathbf{Q})$ -equivariant bundles. We have that  $\tau = \sigma \otimes \nu$ , so we may take the underlying space of  $\tau$  to be  $V_\tau = V_\sigma$  and the action to be  $\tau(g) = \nu(g)\sigma(g) \in \text{End}(V_\sigma)$  for all  $g \in K_{\mathbf{C}}^h$ . Then the map

$$\begin{aligned} G(\mathbf{C}) \times_{Q^-} V_\sigma &\longrightarrow G(\mathbf{C}) \times_{Q^-} V_\tau, \\ (g, w) &\longmapsto (g, \nu(g)^{-1}w) \end{aligned}$$

gives the required isomorphism  $\mathcal{V}_\sigma^\vee \xrightarrow{\sim} \mathcal{V}_\tau^\vee$ . (Note, however, that the Hecke action on the cohomology of  $\mathcal{V}_\sigma$  will depend on  $c$ ; changing the value of  $c$  introduces a corresponding twist by a power of the similitude character in the Hecke action.)

For  $\mu \in X^*(H_{\mathbf{C}})_{K_{\mathbf{C}}^h}^+$  a dominant weight, we let  $\mathcal{V}_\mu$  denote the vector bundle associated to the irreducible  $K_{\mathbf{C}}^h$ -representation  $W_\mu$ . We would like to compare these bundles to the bundles introduced in the proof of Theorem 5.1.

### Definition 5.3

Let  $\mu = (a, b; c) \in X^*(T)_M^+$ . We let  $\mathcal{W}_\mu$  denote the canonical extension  $\mathcal{W}_\mu^{\text{can}}$  in the notation of the proof of Theorem 5.1, and we let  $\mathcal{W}_\mu^{\text{sub}} = \mathcal{W}_\mu(-\infty)$ .

We saw above that, as vector bundles over  $X$ , we have

$$\mathcal{W}_\mu \cong \omega(a, b),$$

though the Hecke action on the cohomology of  $\mathcal{W}_\mu$  will depend on  $c$ .

LEMMA 5.4

Let  $\mu = (a, b; c) \in X^*(T)_M^+$ . Then, over  $X(\mathbf{C})$ , we have

$$\mathcal{W}_{(a,b;c)} \cong \mathcal{V}_{(-b,-a;a+b+2c)},$$

compatibly with Hecke actions on cohomology.

*Proof*

It suffices to prove the isomorphism over  $Y$ . Consider the short exact sequence

$$0 \longrightarrow \mathrm{Lie}_{\mathcal{A}^\vee/Y} \longrightarrow \underline{H}_1^{\mathrm{dR}}(\mathcal{A}/Y) \longrightarrow \mathrm{Lie}_{\mathcal{A}/Y} \longrightarrow 0$$

and the Poincaré duality pairing (see [52, Section 1.2])

$$\langle \cdot, \cdot \rangle : \underline{H}_1^{\mathrm{dR}}(\mathcal{A}/Y) \otimes \underline{H}_1^{\mathrm{dR}}(\mathcal{A}/Y) \longrightarrow \mathcal{O}_Y(1).$$

Expressed in terms of the functor  $\mathcal{W}_\mu$  of Lan and Suh, the short exact sequence becomes

$$0 \longrightarrow \mathcal{W}_{(1,0;0)} \longrightarrow \underline{H}_1^{\mathrm{dR}}(\mathcal{A}/Y) \longrightarrow \mathcal{W}_{(0,-1;1)} \longrightarrow 0$$

and the bundle  $\mathcal{O}_Y(1)$  becomes  $\mathcal{W}_{(0,0;1)}$ . (See [52, Example 1.22].)

Similarly, over  $Y(\mathbf{C})$ , the short exact sequence becomes

$$0 \longrightarrow \mathcal{V}_{(0,-1;1)} \longrightarrow \underline{H}_1^{\mathrm{dR}}(\mathcal{A}/Y) \longrightarrow \mathcal{V}_{(1,0;1)} \longrightarrow 0$$

and  $\mathcal{O}_Y(1)$  is identified with  $\mathcal{V}_{(0,0;2)}$ . This follows from [54, Example III.2.4]. If we take the point  $o \in \check{X}$  to be  $h(i) = J$  in the notation of Section 2.2, then the isotropic subspace corresponds to  $V^-$  and  $V/W$  corresponds to  $V^+$ . As remarked at the end of Section 2.2, we have  $V^- = W_{(0,-1;1)}$ ,  $V^+ = W_{(1,0;1)}$  and the similitude character corresponds to  $W_{(0,0;2)}$ . Note also that the notation  $\mathcal{H}_{\mathrm{dR}}(\mathcal{A})$  of [54, Section I.3] refers to de Rham homology.

It follows that, over  $Y(\mathbf{C})$ , we have  $\mathcal{W}_{(0,0;1)} = \mathcal{V}_{(0,0;2)}$  and  $\mathcal{W}_{(1,0;0)} = \mathcal{V}_{(0,-1;1)}$ . Thus,

$$\begin{aligned} \mathcal{W}_{(a,b;c)} &= (\mathrm{Sym}^{a-b} \otimes \det^b)(\mathcal{W}_{(1,0;0)}) \otimes \mathcal{W}_{(0,0;c)} \\ &= (\mathrm{Sym}^{a-b} \otimes \det^b)(\mathcal{V}_{(0,-1;1)}) \otimes \mathcal{V}_{(0,0;2c)} \\ &= \mathcal{V}_{(-b,-a;a+b+2c)}. \end{aligned}$$

This is compatible with Hecke action on cohomology since all isomorphisms respect the equivariant constructions.  $\square$

The Weyl chambers  $C_0, \dots, C_4 \subset X^*(T) \otimes_{\mathbf{Z}} \mathbf{R} \cong \mathbf{R}^3$  are defined in Section 2.1. We have

$$\begin{aligned} C_0 &= \{(a, b; c) \in \mathbf{R}^3 : a \geq b \geq 0\}, \\ C_1 &= \{(a, b; c) \in \mathbf{R}^3 : a \geq -b \geq 0\}, \\ C_2 &= \{(a, b; c) \in \mathbf{R}^3 : -b \geq a \geq 0\}, \\ C_3 &= \{(a, b; c) \in \mathbf{R}^3 : -b \geq -a \geq 0\}. \end{aligned}$$

**THEOREM 5.5**

Let  $\mu = (a, b; c) \in X^*(T)_M^+$ . Then

$$H^i(X(\mathbf{C}), \mathcal{W}_\mu)_{(2), \text{temp}} = 0$$

for all  $0 \leq i \leq 3$  such that

$$\mu = (a - 1, b - 2; c) \notin C_i.$$

*Proof*

In Section 2.2, we identified  $X^*(H_{\mathbf{C}})$  with  $\mathbf{Z}^3$ . Under the resulting identification of  $X^*(H_{\mathbf{C}}) \otimes_{\mathbf{Z}} \mathbf{R}$  with  $\mathbf{R}^3$ , the chambers  $C_i$  for  $X^*(T)_M^+ \otimes_{\mathbf{Z}} \mathbf{R}$  above correspond to Weyl chambers in  $X^*(H_{\mathbf{C}}) \otimes_{\mathbf{Z}} \mathbf{R}$ . Let  $\sigma = (-b, -a; a + b + 2c)$ , regarded as an element of  $X^*(H_{\mathbf{C}})$ . Then we have seen above that

$$\mathcal{V}_\sigma \cong \mathcal{W}_\mu.$$

Suppose that

$$H^i(X(\mathbf{C}), \mathcal{W}_\mu)_{(2), \text{temp}} = H^i(X(\mathbf{C}), \mathcal{V}_\sigma)_{(2), \text{temp}} \neq 0.$$

Then, by Theorem 5.2, there is some  $\pi = \pi^\infty \otimes \pi_\infty$  in  $\mathcal{A}_{(2)}(G)_{\text{temp}}$  such that

$$H^i(\text{Lie } P^-, K^h; \pi_\infty \otimes V_\sigma) \neq 0.$$

By [37, Theorem 3.5],  $\pi_\infty$  is a discrete series or limit of discrete series. Hence, using the Harish-Chandra parameterization, we may write  $\pi_\infty = \pi(\lambda, C)^* = \pi(-w_0(\lambda), -w_0(C))$  for some Weyl chamber  $C \in \{C_0, \dots, C_3\}$  and a weight  $\lambda \in C \cap (X^*(H_{\mathbf{C}}) + \rho)$ . By [9, Theorem 3.2.1], it follows that

$$\lambda = ((\sigma + \rho)|_{\text{Sp}_4(\mathbf{R})}; -a - b - 2c) = (2 - b, 1 - a; -a - b - 2c)$$

and

$$i = \#(\Phi(C)^+ \cap \Phi_n^+),$$

where  $\Phi(C)^+$  is the system of positive roots determined by the chamber  $C$ . For  $j = 0, \dots, 3$ , we have  $\#(\Phi(C_j)^+ \cap \Phi_n^+) = 3 - j$ . Hence we must have  $C = C_{3-i}$  and  $\lambda \in C_{3-i}$ . However,  $C_{3-i} = -w_0(C_i)$ , so  $-w_0(\lambda) \in C_i$ . We have then

$$\begin{aligned} -w_0(\lambda) &= -w_0(-b + 2, -a + 1; -(a + b + 2c)) \\ &= (a - 1, b - 2; a + b + 2c). \end{aligned}$$

Thus we deduce that  $-w_0(\lambda) = (a - 1, b - 2; a + b + 2c)$  lies in  $C_i$ . This is equivalent to the condition in the statement of the theorem.  $\square$

We also record the following.

**THEOREM 5.6**

Let  $\mu = (a, b; c) \in X^*(T)_M^+$ , let  $w = -(a + b + 2c)$ , and let  $\sigma = (-b, -a; a + b + 2c) = (-b, -a; -w)$ , regarded as an element of  $X^*(H_{\mathbf{C}})$ . Suppose that  $\pi = \pi^\infty \otimes \pi_\infty$  in  $\mathcal{A}_{(2)}(G)$  contributes to  $H^i(X(\mathbf{C}), \mathcal{W}_\mu)_{(2)} \cong H^i(X(\mathbf{C}), \mathcal{V}_\sigma)_{(2)}$ .

- (1) The infinitesimal character of  $\pi_\infty$  is given under the Harish-Chandra isomorphism by

$$\chi_{((- \sigma - \rho)|_{\mathrm{Sp}_4(\mathbf{R})}; -w)} = \chi_{(a-1, b-2; -w)}.$$

- (2) Let  $\tilde{\pi}_\infty$  denote the transfer of  $\pi_\infty$  to  $\mathrm{GL}_4(\mathbf{R})$ . Then the infinitesimal character of  $\tilde{\pi}_\infty$  is given under the Harish-Chandra isomorphism by  $\chi_\tau$  where

$$\tau = \left( \frac{a + b - 3 - w}{2}, \frac{a - b + 1 - w}{2}, \frac{-a + b - 1 - w}{2}, \frac{-a - b + 3 - w}{2} \right).$$

- (3) If, furthermore,  $\pi_\infty$  is tempered, then  $\pi_\infty$  is a discrete series or limit of discrete series representation, and is given under the Harish-Chandra parameterization by

$$\pi_\infty \cong \pi((a - 1, b - 2; -w), C_i).$$

*Proof*

For the first part, we have that

$$H^i(\mathrm{Lie} P^-, K^h; \pi_\infty \otimes V_\sigma) \neq 0.$$

It follows from [9, Theorem 3.2.1] that the infinitesimal character of  $\pi_\infty$  is equal to  $\chi_{((- \sigma - \rho)|_{\mathrm{Sp}_4(\mathbf{R})}; -w)}$ . The second part can be inferred from [64, Section 2.1.2]. The last part was established in the proof of Theorem 5.5.  $\square$

*Definition 5.7*

A weight  $\mu = (a, b; c) \in X^*(T)_M^\dagger$  such that  $(a - 1, b - 2; c)$  lies in the interior of a unique Weyl chamber  $C_i$  is said to be a *discrete series weight* or a *regular weight*. If  $\mu - w_0(\rho)$  lies in the intersection of exactly two of Weyl chambers  $C_i$ , then we say that it is a *limit of discrete series weight* or a *nonregular weight*.

From the explicit description of the Weyl chambers  $C_i$  above, we see that the limit of discrete series weights thus come in three families:

$$\mu = (a, 2; c) \quad \text{with } a \in \mathbf{Z}_{\geq 2},$$

$$\mu = (a, 3 - a; c) \quad \text{with } a \in \mathbf{Z}_{\geq 2},$$

$$\mu = (1, b; c) \quad \text{with } b \in \mathbf{Z}_{\leq 1}.$$

Note that for the corresponding families of vector bundles  $\mathcal{W}_\mu = \omega(a, b)$ , the first and third are interchanged under the Serre duality map  $\omega(a, b) \mapsto \omega(a, b)^\vee \otimes \det \Omega_X^1 \cong \omega(3 - b, 3 - a)(-\infty)$ , while the second family is stable under this operation. (Up to interchanging the canonical and subcanonical extensions, of course.) The preceding theorem implies that for all  $a \geq 2$ , we have

$$H^i(X(\mathbf{C}), \omega(a, 2))_{(2), \text{temp}} = 0 \quad \text{for } i = 2, 3,$$

$$H^i(X(\mathbf{C}), \omega(a, 3 - a))_{(2), \text{temp}} = 0 \quad \text{for } i = 0, 3.$$

(Technically, we should normalize the Hecke action on the cohomology of  $\omega(a, b)$  before we adjoin the subscripts (2) or temp; see Section 5.5 below.) From the result of Lan and Suh, we deduce the following characteristic  $p$  analogue of these vanishing results for limit of discrete series weights.

**COROLLARY 5.8**

(1) For  $4 \leq a \leq p$ , we have

$$H^i(X, \omega(a, 2)(-\infty)_k) = 0$$

for  $i = 2, 3$ .

(2) For  $3 \leq a \leq (p + 1)/2$ , we have

$$H^0(X, \omega(a, 3 - a)_k) = H^3(X, \omega(a, 3 - a)(-\infty)_k) = 0.$$

*Proof*

The vanishing results for the subcanonical extensions  $\omega(*, *)(-\infty)$  follow directly from Theorem 5.1. The fact that

$$H^0(X, \omega(a, 3-a)_k) = 0$$

in the second part then follows from Serre duality since

$$\begin{aligned} \omega(a, 3-a)^\vee \otimes \det \Omega_{X/\mathcal{O}}^1 &= \omega(a-3, -a) \otimes \omega(3, 3)(-\infty) \\ &= \omega(a, 3-a)(-\infty). \end{aligned} \quad \square$$

#### 5.4. Torsion classes

It seems natural to ask whether one can (explicitly or otherwise) construct classes in  $H^0(X, \omega(2, 2))$  which do not lift to characteristic 0. Let us recall what happens for classical modular forms of weight 1.

Suppose that  $X_1(N)$  denotes (for this paragraph) the classical modular curve. A non-Eisenstein Hecke eigenclass in  $H^0(X_1(N), \omega_k)$  gives rise to an irreducible Galois representation  $\bar{r} : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(k)$ . Suppose that the image of  $\bar{r}$  contains  $\mathrm{SL}_2(k')$  for some  $\#k' > 5$ . Such a representation cannot be the mod- $p$  reduction of a representation with image isomorphic to some subgroup of  $\mathrm{GL}_2(\mathbf{C})$ , and thus by [23], the corresponding mod- $p$  class does not lift to characteristic 0. (Explicit examples were first found by Mestre for  $\#k = 8$  and  $N = 1429$ .) A slightly different example can be given as follows. Suppose that  $\Gamma = \Gamma_1(N) \cap \Gamma_0(x)$ . Consider a non-Eisenstein Hecke eigenclass in  $H^1(X(\Gamma), \omega_k)$  which is new of level  $x$ . Then the restriction of  $\bar{r}$  to  $I_x$  is rank 2 unipotent. Such a class cannot lift to characteristic 0 at minimal level, because otherwise (by [23] again) the corresponding representation  $\rho$  would simultaneously have finite image and yet  $\rho|_{I_x}$  would be unipotent and hence infinite. Note that (unlike in the first example) it may well be possible to lift  $\bar{r}$  to characteristic 0 at some nonminimal level. Examples of the second kind have a natural analogue in the Siegel context.

Suppose that  $\bar{r}$  has type **U3** at  $x$ . If  $r$  is any minimal lift of  $\bar{r}$ , then the image of  $I_x$  under  $r$  will be rank 3 unipotent. This will also be true for the restriction of  $r$  to any finite extension of  $\mathbf{Q}_x$ . Yet, by a theorem of Grothendieck ([36, Exposé 9]), the image of inertia of a semistable abelian variety is rank 2 unipotent; that is, it satisfies  $(\sigma - 1)^2 = 0$ . It follows that  $r$  cannot contribute to a motive associated to an abelian variety. Conjecturally, Siegel modular eigenforms of weight  $(2, 2)$  should be associated to abelian varieties  $M/\mathbf{Q}$  of dimension  $2n$  equipped with an injection  $E \rightarrow \mathrm{End}_{\mathbf{Q}}(M) \otimes \mathbf{Q}$  for some totally real field  $E$  of degree  $n$ . This suggests that such representations  $\bar{r}$  do not admit minimal lifts to characteristic 0 when  $\sigma = (2, 2)$ . It would be interesting to produce an explicit example of such a modular representation. Recall that there is an exceptional isomorphism  $S_6 \simeq \mathrm{GSp}_4(\mathbf{F}_2)$  coming from identifying the Galois group of  $\mathcal{A}_2[2]$  over  $\mathcal{A}_2$  with either the symmetries of the 2-torsion points on the universal abelian surface or the action of  $S_6$  on the (generically) six Weierstrass points (see [8]). The unipotent element  $\sigma \in \mathrm{GSp}_4(\mathbf{F}_2)$  such that

$(\sigma - 1)^2 \neq 0$  has conjugacy class  $(1, 2, 3, 4)(5, 6) \in S_6$  (this class is preserved by the exotic automorphism of  $S_6$ ). In particular, if  $K/\mathbf{Q}$  is a sextic field with Galois closure  $G \subset S_6$  containing  $(1, 2, 3, 4)(5, 6)$  and acting irreducibly on  $\mathbf{F}_2^4$ , and  $p$  is an odd prime such that  $p = p^4 q^2$ , then  $\bar{r} : \text{Gal}(K/\mathbf{Q}) \simeq \text{GSp}_4(\mathbf{F}_2)$  should give rise to such a representation. Here is an explicit example coming from a slight variation of this argument. Suppose that  $A$  is the abelian surface corresponding to the Jacobian of the curve

$$y^2 = x^5 - 2x^4 + 6x^3 - 8x^2 + 4x - 4,$$

which has good reduction outside  $3 \cdot 5 \cdot 19$ . The representation  $\bar{r} : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\mathbf{F}_2)$  has image  $S_5 \subset S_6$ , and the image of inertia at 5 is conjugate to  $(1, 2, 3, 4)(5, 6)$ . Hence  $\bar{r}$  should give rise to a mod-2 torsion class with trivial level structure outside  $3 \cdot 5 \cdot 19$ , and the following level structure at these primes:

- (1) Iwahori level structure at  $p = 5$ ,
- (2) paramodular level structure at  $p = 3$  and 19.

Note that this conjectural torsion class *does* conjecturally lift to characteristic 0 at some level since one expects that  $A$  is modular. (The conductor of  $A$  is  $3 \cdot 5^3 \cdot 19$ .)

Common to both examples is the nonexistence of automorphic representations  $\pi$  (associated to either classical modular forms of weight 1 or Siegel modular forms of weight  $(2, 2)$ ) such that  $\pi_x$  is the Steinberg representation. For classical modular forms, the nonexistence of such  $\pi$  follows from a consideration of the corresponding Galois representations, an argument which does not obviously generalize to the Siegel case (since one does not know how to attach an abelian variety to such a form). However, the following argument (due to Kevin Buzzard) generalizes nicely

#### THEOREM 5.9

*If  $\pi$  is a cuspidal automorphic representation associated to a Siegel modular form of weight  $(2, 2)$ , then  $\pi_x$  is not the Steinberg representation for any  $p$ .*

#### *Proof*

In weights  $(j, k)$  with  $j \geq k \geq 2$ , the corresponding Frobenius eigenvalues of the Weil–Deligne representation associated to a Steinberg representation  $\pi_x$  are

$$\{x^{(w+3)/2}, x^{(w+1)/2}, x^{(w-1)/2}, x^{(w-3)/2}\},$$

where  $w = j + k - 3$ . Moreover, the corresponding eigenvalue of  $U_{x,1}$  is  $x^{(w-3)/2}$ . In particular, if  $j = k = 2$ , then  $w = 1$  and the corresponding eigenvalue of  $U_{x,1}$  is  $x^{-1}$ , contradicting the integrality of Hecke eigenvalues (which is a consequence of the integrality of the  $q$ -expansion).  $\square$



### 5.5. Hecke operators

For simplicity, we denote the schemes  $X_K$  and  $X_{K_i(Q)}$  of Section 5.2 by  $X$  and  $X_i(Q)$ , respectively. Let  $M$  denote an  $\mathcal{O}$ -module.

Let  $x$  be a rational prime. We define matrices

$$\beta_{x,0} = \begin{pmatrix} x & 0 & 0 & 0 \\ 0 & x & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x \end{pmatrix}, \quad \beta_{x,1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x \end{pmatrix},$$

$$\beta_{x,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x & 0 & 0 \\ 0 & 0 & x & 0 \\ 0 & 0 & 0 & x^2 \end{pmatrix}$$

and regard them as elements of  $\mathrm{GSp}_4(\mathbf{Q}_x)$ . If  $x \notin S$  (resp.,  $x \notin S \cup Q$ ), then we will consider the Hecke operators  $T_{x,i} = [K\beta_{x,i}K]$  (resp.,  $T_{x,i} = [K_i(Q)\beta_{x,i}K_i(Q)]$ ) acting on each of the spaces

$$H^n(X, \omega(a, b)_M) \quad (\text{resp., } H^n(X_i(Q), \omega(a, b)_M))$$

as in [63, Section 1.1.6] or [68, Section 8]. We also denote  $T_{x,0}$  by  $S_x$ . The definition of Hecke operators given in [63] or [68] applies when  $x \neq p$  or when  $p$  is invertible on  $M$ . The remaining cases when  $x = p$  require more care. In Lemma 8.8 below we show that  $T_{p,1}$  and  $Q_{p,2} := (pT_{p,2} + (p + p^3)S_p)p^{2-b}$  exist as operators in cohomological degree  $n = 0$  over  $M = K/\mathcal{O}$ .

Similarly, if  $x \in Q$ , then we have operators  $U_{x,i} = [K_i(Q)\beta_{x,i}K_i(Q)]$  on  $H^n(X_i(Q), \omega(a, b)_M)$ . As in Section 5.2, the map  $X_1(Q) \rightarrow X_0(Q)$  is Galois with Galois group  $\Delta_Q := \prod_{x \in Q} (\mathbf{Z}/x)^\times$ . This gives rise to an action of  $\Delta_Q$  on  $H^n(X_1(Q), \omega(a, b)_M)$ . For each  $u \in \Delta_Q$ , we denote the corresponding operator on  $H^n(X_1(Q), \omega(a, b)_M)$  by  $\langle u \rangle$ .

Finally, we will also exploit Hecke operators of a slightly different flavor, which we denote by  $U_{p,1}$  and  $U_{p,2}$ , respectively. In the context of this paper, they may be considered formal operators on  $q$ -expansions. (They can also be interpreted more classically as Hecke operators with level structure at  $p$ .) Their key property is that the operators  $T_{p,1}$  and  $T_{p,2}/p^{k+j-6}$  act by  $U_{p,1}$  and  $U_{p,2}$  for large enough weights, including  $(j, k)$  plus any nontrivial multiple of  $(p - 1, p - 1)$  for  $j \geq k \geq 2$ . Their explicit definition is given in Lemmas 8.3 and 8.4.

#### Remark 5.10

We note that our definition of the Hecke action is the “natural” one twisted by  $v^{-3}$  (see [63, Section 1.1.6(a)]). We saw in the proof of Theorem 5.1 that, for the natural action,

there is an isomorphism  $\omega(a, b) \cong \mathcal{W}_{(a,b;-a-b)}$ , and hence over  $\mathbf{C}$ , an isomorphism  $\omega(a, b) \cong \mathcal{V}_{(-b,-a;-a-b)}$ . Under our normalization of the Hecke action on  $\omega(a, b)$ , we therefore have  $\omega(a, b) \cong \mathcal{W}_\mu$  and, over  $\mathbf{C}$ ,  $\omega(a, b) \cong \mathcal{V}_\sigma$  where we take

$$\mu = (a, b; 3 - a - b) \quad \text{and} \quad \sigma = (-b, -a; 6 - a - b).$$

*Remark 5.11*

In view of the previous remark, we will identify the set  $\mathbf{Z}^{2,+} := \{(a, b) \in \mathbf{Z}^2 : a \geq b\}$  with the subset  $(a, b; 3 - a - b)$  of  $X^*(T)_M^+$ . Thus it makes sense to speak of  $\mu = (a, b) \in X^*(T)_M^+$ .

*Remark 5.12*

Let  $\mu = (a, b) \in X^*(T)_M^+$ , and let  $w = a + b - 6$ . For  $x \in \mathbf{Z}$ , we can similarly define a Hecke operator associated to  $[K \text{diag}(x, x, x, x)K]$  on the cohomology of  $\omega(a, b)$ : this operator acts as  $x^w = x^{a+b-6}$ . Now, suppose that  $\pi = \pi^\infty \otimes \pi_\infty$  in  $\mathcal{A}_{(2)}(G)$  contributes to

$$H^i(X(\mathbf{C}), \omega(a, b))_{(2)} \cong H^i(X(\mathbf{C}), \mathcal{W}_\mu)_{(2)} \cong H^i(X(\mathbf{C}), \mathcal{V}_\sigma)_{(2)},$$

where  $\sigma = (-b, -a; -w)$ . It follows that the central character of  $\pi_\infty$  is given by

$$x \mapsto x^{-w}.$$

Furthermore, by Proposition 5.6, the transfer of  $\pi_\infty$  to  $\text{GL}_4(\mathbf{R})$  has infinitesimal character  $\chi_\tau$ , where

$$\tau = (0, -(b-2), -(a-1), -(a+b-3)) + 3/2(1, 1, 1, 1).$$

We now introduce some Hecke algebras. We note that in the following definition, we work over  $K/\mathcal{O}$  rather than  $\mathcal{O}$ .

*Definition 5.13*

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ .

- (1) The anemic Hecke algebra

$$\mathbf{T}_\mu^{\text{an}}(\mathcal{Q}) \subset \text{End}_{\mathcal{O}}(H^0(X_1(\mathcal{Q}), \omega(a, b)(-\infty)_{K/\mathcal{O}}))$$

is the  $\mathcal{O}$ -algebra generated by the operators  $T_{x,i}$  for  $x \notin S \cup \mathcal{Q} \cup \{p\}$ .

- (2) Similarly, we let  $\mathbf{T}_\mu(\mathcal{Q})$  be the algebra generated over  $\mathbf{T}_\mu^{\text{an}}(\mathcal{Q})$  by the operators  $U_{x,i}$  for  $x \in \mathcal{Q}$  and  $\langle u \rangle$  for  $u \in \Delta_{\mathcal{Q}}$ . When  $\mathcal{Q} = \emptyset$ , we have  $\mathbf{T}_\mu^{\text{an}}(\emptyset) = \mathbf{T}_\mu(\emptyset)$  and we denote this algebra by  $\mathbf{T}_\mu$ .

- (3) Finally,  $\widetilde{\mathbf{T}}_\mu(Q)$  denotes the  $\mathbf{T}_\mu(Q)$ -algebra generated by the operators  $T_{p,1}$  and  $Q_{p,2} = (pT_{p,2} + (p + p^3)S_p)p^{2-b}$ . (The existence of these operators is established in Lemma 8.8.) If  $Q = \emptyset$ , then we denote  $\widetilde{\mathbf{T}}_\mu(\emptyset)$  by  $\widetilde{\mathbf{T}}_\mu$ .

Note that the algebras  $\mathbf{T}_\mu^{\text{an}}(Q) \subset \mathbf{T}_\mu(Q) \subset \widetilde{\mathbf{T}}_\mu(Q)$  preserve the subspace

$$H^0(X_0(Q), \omega(a, b)(-\infty)_{K/\emptyset}) \subset H^0(X_1(Q), \omega(a, b)(-\infty)_{K/\emptyset}).$$

We will also need to consider ordinary Hecke algebras. Let  $e = \lim_{\rightarrow n} (T_{p,1}Q_{p,2})^{n!}$  denote the ordinary idempotent associated to the Hecke operators  $T_{p,1}$  and  $Q_{p,2}$ . (We will only consider this operator in contexts where the direct limit makes sense.) We define

$$H^0(X_0(Q), \omega(a, b)(-\infty)_M)^{\text{ord}} = eH^0(X_0(Q), \omega(a, b)(-\infty)_M)$$

for  $M = \emptyset, \emptyset/\varpi^m$  or  $M = K/\emptyset$ . We thus have

$$\begin{aligned} & H^0(X_0(Q), \omega(a, b)(-\infty)_M) \\ &= H^0(X_0(Q), \omega(a, b)(-\infty)_M)^{\text{ord}} \oplus (1 - e)H^0(X_0(Q), \omega(a, b)(-\infty)_M) \end{aligned}$$

for such  $M$ .

#### Definition 5.14

Let  $\mu = (a, b)$  with  $a \geq b \geq 2$ . We define the ordinary Hecke algebras  $\mathbf{T}_\mu^{\text{an}}(Q)^{\text{ord}}$  (resp.,  $\mathbf{T}_\mu(Q)^{\text{ord}}, \widetilde{\mathbf{T}}_\mu(Q)^{\text{ord}}$ ) to be the image of  $\mathbf{T}_\mu^{\text{an}}(Q)$  (resp.,  $\mathbf{T}_\mu(Q), \widetilde{\mathbf{T}}_\mu(Q)$ ) in

$$\text{End}_\emptyset(H^0(X_0(Q), \omega(a, b)(-\infty)_{K/\emptyset})^{\text{ord}}).$$

## 6. Galois representations associated to modular forms

As in Section 5.2, let  $S$  and  $Q$  be finite sets of primes of  $\mathbf{Q}$  which are disjoint and do not contain  $p$ . We allow the possibility that  $Q = \emptyset$ . We let  $K$  and  $K_i(Q)$  be open compact subgroups of  $\text{GSp}_4(\mathbb{A}^\infty)$  as in Section 5.2, and we let  $X = X_K$  and  $X_i(Q) = X_{K_i(Q)}$  be the corresponding Siegel 3-folds, defined over  $\emptyset$ .

### 6.1. The Hasse invariant

We begin with a definition.

#### Definition 6.1

Let  $h \in H^0(X, \omega_k^{p-1})$  be the Hasse invariant, and let  $A \in H^0(X, \omega^r(p-1))$  be a lift of  $h^r$ , for some  $r > 0$  which we fix for the rest of this section.

The existence of such a lift  $A$  follows from the Koecher principle and the ampleness of  $\omega$  on the minimal compactification of  $X$ .

LEMMA 6.2

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ . Then

(1) multiplication by  $h$  defines an injection

$$h : H^0(X_1(Q), \omega(a, b)_k) \hookrightarrow H^0(X_1(Q), \omega(a + (p-1), b + (p-1))_k)$$

which is equivariant for the Hecke operators  $T_{x,i}$  for each  $x \notin S \cup Q \cup \{p\}$  and the operators  $U_{x,i}$  for  $x \in Q$ , and

(2) if  $b \geq 3$ , then this map is also equivariant for the operators  $T_{p,1}$  and  $Q_{p,2}$ .

*Proof*

It is well known that multiplication by  $h$  is injective and commutes with Hecke operators away from  $p$ . We may thus assume that  $b \geq 3$ . It is shown in [57, Section A.3] and [68, Lemme 8.7] that multiplication by  $h$  commutes with the operators  $U_{p,1}$  and  $U_{p,2}$ . Since  $b \geq 3$ , [68, Lemme 8.5] implies that  $T_{p,1} \equiv U_{p,1} \pmod{p}$  and  $p^{3-b}T_{p,2} \equiv U_{p,2} \pmod{p}$ . It follows that  $T_{p,1}$  and  $Q_{p,2} = p^{3-b}T_{p,2} + (1 + p^2)p^{3-b}S_p$  also commute with  $h$ .  $\square$

Suppose that  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ . By the proof of [57, Théorème 6.2], there exists an integer  $N(\mu)$  as in the following definition.

*Definition 6.3*

Let  $N(\mu)$  be an integer such that for all  $t \geq N(\mu)$ ,  $i > 0$ , and  $Z \in \{X, X_0(Q), X_1(Q)\}$ , the cohomology group

$$H^i(Z, \omega(a+t, b+t)(-\infty)_k)$$

vanishes.

Note that for such  $t \geq N(\mu)$ , the maps

$$\begin{aligned} H^0(X, \omega(a+t, b+t)(-\infty)) &\rightarrow H^0(X, \omega(a+t, b+t)(-\infty)_k), \\ H^0(X, \omega(a+t, b+t)(-\infty)_K) &\rightarrow H^0(X, \omega(a+t, b+t)(-\infty)_{K/\mathcal{O}}) \end{aligned}$$

are both surjective. The same is true over  $X_0(Q)$  and  $X_1(Q)$ .

LEMMA 6.4

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$  and let  $m > 0$ . There exists an integer  $s > 0$  such that, if we set  $t = rs(p-1)$ , then

- (1)  $t \geq N(\mu)$ , and  
 (2) multiplication by  $A^s$  defines an injection

$$H^0(X_1(Q), \omega(a, b)_{\mathfrak{o}/\mathfrak{w}^m}) \hookrightarrow H^0(X_1(Q), \omega(a+t, b+t)_{\mathfrak{o}/\mathfrak{w}^m})$$

which is equivariant for the Hecke operators  $T_{x,i}$  for each  $x \notin S \cup Q \cup \{p\}$  and the operators  $U_{x,i}$  for each  $x \in Q$ .

*Proof*

The second property holds as long as  $p^{m-1} \mid s$  (see [32, Theorem 6.2.1]), so it suffices to take  $s$  equal to any integer greater than  $N(\mu)/r(p-1)$  and divisible by  $p^{m-1}$ .  $\square$

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ . Recall that the Hecke algebras

$$\mathbf{T}_\mu^{\text{an}}(Q) \subset \mathbf{T}_\mu(Q) \subset \widetilde{\mathbf{T}}_\mu(Q) \subset \text{End}_{\mathfrak{o}}(H^0(X_1(Q), \omega(a, b)(-\infty)_{K/\mathfrak{o}}))$$

were defined in Definition 5.13.

*Remark 6.5*

For  $\mu = (a, b)$  with  $a \geq b \geq 2$  and each  $m > 0$ , we have

$$H^0(X_1(Q), \omega(a, b)(-\infty)_{\mathfrak{o}/\mathfrak{w}^m}) \cong H^0(X_1(Q), \omega(a, b)(-\infty)_{K/\mathfrak{o}})[\mathfrak{w}^m].$$

Let  $I_{\mu,m}$  (resp.,  $\widetilde{I}_{\mu,m}$ ) denote the annihilator of the former space in  $\mathbf{T}_\mu(Q)$  (resp.,  $\widetilde{\mathbf{T}}_\mu(Q)$ ). If  $s$  and  $t$  are as in Lemma 6.4, then multiplication by  $A^s$  induces a surjective map

$$\mathbf{T}_{\mu'}(Q) \twoheadrightarrow \mathbf{T}_\mu(Q)/I_{\mu,m},$$

where  $\mu' = \mu + (t, t)$ . In particular, any maximal ideal  $\mathfrak{m}$  of  $\mathbf{T}_\mu(Q)$  pulls back under this map to a maximal ideal of  $\mathbf{T}_{\mu'}(Q)$  which we will also denote by  $\mathfrak{m}$ .

Similarly, Lemma 6.2 induces a map

$$\mathbf{T}_{\mu'}(Q) \twoheadrightarrow \mathbf{T}_\mu(Q)/I_{\mu,1}$$

where  $\mu' = \mu + (p-1, p-1)$  and, if  $b \geq 3$ , this extends to a map

$$\widetilde{\mathbf{T}}_{\mu'}(Q) \twoheadrightarrow \widetilde{\mathbf{T}}_\mu(Q)/\widetilde{I}_{\mu,1}.$$

## 6.2. Preliminaries on Galois representations

We now turn our attention to Galois representations.

## PROPOSITION 6.6

Let  $\mu = (a, b) \in X^*(T)_M^+$ , and let  $w = a + b - 6$ . There is a continuous character

$$\chi_\mu : G_{\mathbf{Q}} \rightarrow \mathbf{T}_\mu^{\text{an}}(Q)^\times$$

such that

- (1)  $\chi_\mu \mid G_{\mathbf{Q}_p}$  is crystalline with Hodge–Tate weight  $w$ ,
  - (2) for all  $x \notin S \cup Q \cup \{p\}$ ,  $\chi_\mu$  is unramified at  $x$  and  $\chi_\mu(\text{Frob}_x) = S_x$ ;
- in particular,

$$\chi_\mu = \chi_{\mu,0} \epsilon^{-w}$$

for some finite-order character  $\chi_{\mu,0} : G_{\mathbf{Q}} \rightarrow \tilde{\mathbf{T}}_\mu(Q)^\times$ .

*Proof*

This follows from the proof of [65, Proposition 4], noting that we have twisted the Hecke action by  $v^{-3}$  (see Remark 5.12).  $\square$

## Definition 6.7

For a prime  $x$ , we introduce the Hecke polynomial

$$Q_x(T) = X^4 - T_{x,1}X^3 + (xT_{x,2} + (x^3 + x)S_x)X^2 - x^3S_xT_{x,1}X + x^6S_x^2.$$

If a modular form  $f$  is an eigenform for a collection of Hecke operators  $T$ , we denote by  $\lambda_f$  the map such that  $Tf = \lambda_f(T)f$  for each  $T$ . In particular, if  $f$  is an eigenform for the operators  $T_{x,i}$  at  $x$ , then we can specialize the polynomial  $Q_x(T)$  at  $f$  to get  $\lambda_f(Q_x(T))$ .

## PROPOSITION 6.8

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 3$ . Let  $w = a + b - 6$ , and let  $\mathbf{w} = w + 3 = a + b - 3$ . Let

$$f \in H^0(X_1(Q), \omega(a, b)(-\infty))$$

be a cuspidal eigenform for the operators  $T_{x,i}$  for all  $x \notin Q \cup S$  and  $i = 0, 1, 2$ . Then there is a continuous semisimple representation

$$r_f : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(K')$$

defined over a finite extension  $K'/K$  such that:

- (1) The similitude character  $v \circ r_f$  is given by

$$v \circ r_f = \lambda_f \circ \chi_\mu \epsilon^{-3} = \lambda_f \circ \chi_{\mu,0} \epsilon^{-\mathbf{w}}.$$

- (2) *The representation  $r_f$  is unramified at primes  $x \notin Q \cup S \cup \{p\}$ , and at such primes, the characteristic polynomial of  $r_f(\text{Frob}_x)$  is given by*

$$\det(X - r_f(\text{Frob}_x)) = \lambda_f(Q_x(X)).$$

- (3) *The restriction  $r_f | G_{\mathbf{Q}_p}$  is crystalline with Hodge–Tate weights  $\mathbf{w}$ ,  $(a - 1)$ ,  $(b - 2)$ ,  $0$ . If, in addition,  $f$  is an eigenvalue of the Hecke operators at  $p$ , then the characteristic polynomial of  $\Phi$  on  $D_{\text{cris}}(r_f | G_{\mathbf{Q}_p})$  is  $\lambda_f(Q_p(X))$ .*
- (4) *Suppose that  $f$  is ordinary in the sense that it is an eigenform for  $T_{p,1}$  and  $Q_{p,2}$  with eigenvalues being  $p$ -adic units. Then  $Q_p(X)$  has distinct eigenvalues  $\alpha_p, \beta_p, \gamma_p, \delta_p$  with  $p$ -adic valuations  $0, b - 2, a - 1, \mathbf{w}$ , respectively. Furthermore,  $r_f | G_{\mathbf{Q}_p}$  is conjugate in  $\text{GSp}_4(K')$  to a representation of the form*

$$\begin{pmatrix} \lambda(\alpha_p) & & * & & * & & * \\ 0 & \epsilon^{-(b-2)} \cdot \lambda(p^{-(b-2)}\beta_p) & & & * & & * \\ 0 & 0 & & \epsilon^{-(a-1)} \cdot \lambda(p^{-(a-1)}\gamma_p) & & & * \\ 0 & 0 & & 0 & & \epsilon^{-\mathbf{w}} \cdot \lambda(p^{-\mathbf{w}}\delta_p) & * \end{pmatrix}.$$

- (5) *If  $r_f$  is absolutely irreducible, then it satisfies local-global compatibility at all primes.*

*Proof*

The existence of  $r_f$  follows from the work of Taylor, Laumon, and Weissauer. Some of the finer properties are due to Urban, Genestier, Tilouine, Gan, Takeda, Sorensen, and Mok. Fix an embedding  $\iota : K \hookrightarrow \mathbf{C}$ , and let  $\pi$  be a cuspidal automorphic representation of  $\text{GSp}_4(\mathbb{A}_{\mathbf{Q}})$  which contributes to the  $f$ -part of  $H^0(X_1(Q), \omega(a, b)(-\infty)_{\mathbf{C}})$  under the isomorphism of Theorem 5.2(1) (with  $\sigma = (-b, -a; 6 - a - b)$ , as in Remark 5.10).

We take  $r_f : G_{\mathbf{Q}} \rightarrow \text{GL}_4(\overline{K})$  to be the representation  $R_p$  of [55, Theorem 3.5] associated to  $\pi$ . When  $\pi$  is simple, generic in the terminology of [55], the representation can be conjugated to take values in  $\text{GSp}_4(\overline{K})$ , by the main theorem of [6]. In the remaining cases, the representation  $R_p$  is reducible and can easily be seen to be symplectic. The usual Baire category argument implies that  $r_f$  can be defined over a finite extension of  $K$ . Thus in all cases, we may take  $r_f : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(K')$ . Parts (1)–(5) follow from the statement of [55, Theorem 3.5].  $\square$

LEMMA 6.9

*Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ , and let  $\mathfrak{m}$  be a maximal ideal of  $\mathbf{T}_{\mu}^{\text{an}}(Q)$ . Then there is a continuous semisimple representation*

$$\overline{r}_{\mathfrak{m}} : G_{\mathbf{Q}} \rightarrow \text{GL}_4(\mathbf{T}_{\mu}^{\text{an}}(Q)/\mathfrak{m})$$

such that for each  $x \notin S \cup Q \cup \{p\}$ , the restriction  $\bar{r}_m|_{G_{\mathbf{Q}_x}}$  is unramified and  $\bar{r}_m(\text{Frob}_x)$  has characteristic polynomial  $Q_x(X)$ .

If  $\bar{r}_m$  is absolutely irreducible, then the representation  $\bar{r}_m$  preserves a symplectic pairing and hence, after conjugation, we have a representation

$$\bar{r}_m : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\mathbf{T}_{\mu}^{\text{an}}(Q)/\mathfrak{m}).$$

*Proof*

Choose an integer  $s$  as in Lemma 6.4 with  $m$  taken to equal 1, and let  $t = rs(p-1)$ . Let  $f \in H^0(X_1(Q), \omega(a+t, b+t)(-\infty)) \otimes \bar{K}$  be an eigenform for  $\tilde{\mathbf{T}}_{\mu'}(Q)_{\mathfrak{m}}$ . Let  $r_f$  be the Galois representation associated to  $f$  by Proposition 6.8, and take  $\bar{r}_m$  to be the semisimplification of a reduction of  $r_f$  to characteristic  $p$ . The resulting representation is defined over the algebraic closure of  $\mathbf{T}_{\mu}^{\text{an}}(Q)/\mathfrak{m}$ , but by the argument of [20, Proposition 3.4.2], we see that after conjugation, it may be defined over  $\mathbf{T}_{\mu}^{\text{an}}(Q)/\mathfrak{m}$ .

For the last part, let  $\pi$  be the transfer to  $\text{GL}_4$  (given by [3]) of the automorphic representation generated by  $f$ . Then  $\pi$  descends to an automorphic representation  $\Pi$  of a unitary group over  $\mathbf{Q}$ . The family of  $\ell$ -adic Galois representations associated to  $\Pi$  is the same as that associated to  $f$ . Thus, [7, Theorem 1.2] and the fact that  $\bar{r}_m$  is absolutely irreducible imply that  $r_f$  is symplectic. The same is then true of  $\bar{r}_m$  (by absolute irreducibility).  $\square$

*Remark 6.10*

By the same argument, the previous result holds if we replace  $\mathbf{T}_{\mu}^{\text{an}}(Q)$  by  $\mathbf{T}_{\mu}(Q)$  or  $\tilde{\mathbf{T}}_{\mu}(Q)$ .

*Definition 6.11*

We say that  $\mathfrak{m}$  is *non-Eisenstein* if the representation  $\bar{r}_m$  is absolutely irreducible.

### 6.3. Galois representations in cohomological weights

Let  $\bar{r} : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(k)$  be a representation as in Section 4. By Assumption 4.2 and Chebotarev, there exist infinitely many primes  $q$  such that no pair of eigenvalues of  $\bar{r}(\text{Frob}_q)$  have ratio  $q \pmod{p}$  and  $q \not\equiv 1 \pmod{p}$ . Choose any such  $q$  which is disjoint to  $p$  and all primes of bad reduction of  $\bar{r}$ . We take  $S = S(\bar{r}) \cup \{q\}$  and  $Q$  a possibly empty set of primes disjoint from  $S \cup \{p\}$ . We define a compact open subgroup  $K = \prod_x K_x$  of  $\text{GSp}_4(\mathbb{A}^{\infty})$  as follows:

- (1) If  $x = p$  or  $\bar{r}$  is unramified at  $x$  and  $x \neq q$ , then  $K_x = \text{GSp}_4(\mathbf{Z}_x)$ .
- (2) If  $x$  is of type **U3**, then  $K_x = I(x)$ , where  $I(x)$  is the Iwahori subgroup.
- (3) If  $x$  is of type **U2**, then  $K_x = \Pi(x)$ , where  $\Pi(x)$  is the Klingen parahoric.
- (4) If  $x$  is of type **U1**, then  $K_x = K(x)$ , where  $K(x)$  is the paramodular group at  $x$ .



- (5) If  $x$  is of type **P**, then  $K_x = \Pi(x)^+$  (and  $x - 1$  is prime to  $p$ ).
- (6) If  $x$  is of type **H**, then  $K_x$  is the full congruence subgroup of level  $x$ .
- (7) If  $x = q$ , then  $K_x$  is the full congruence subgroup of level  $x$ .

We then let  $X = X_K$  and  $X_i(Q) = X_{K_i(Q)}$  as in Section 5.2.

Let  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 3$  be a *regular weight*, and let  $\mathfrak{m}_\emptyset$  be a maximal ideal of  $\mathbf{T}_\mu^{\text{ord}}$  (the ordinary Hecke algebra with  $Q = \emptyset$ ) with residue field  $k$ . Then  $\mathfrak{m}_\emptyset$  pulls back to an ideal of  $\mathbf{T}_\mu^{\text{an}}(Q)^{\text{ord}}$  which in turn pushes forward to an ideal of  $\mathbf{T}_\mu(Q)^{\text{ord}}$ . We denote both of these ideals by  $\mathfrak{m}_\emptyset$ , in a slight abuse of notation. The ideal  $\mathfrak{m}_\emptyset \subset \mathbf{T}_\mu^{\text{an}}(Q)^{\text{ord}}$  is maximal, but  $\mathfrak{m}_\emptyset \subset \mathbf{T}_\mu(Q)^{\text{ord}}$  need not be maximal—there may be multiple maximal ideals  $\mathfrak{m}$  of  $\mathbf{T}_\mu(Q)^{\text{ord}}$  that contain it. We make the following assumption.

*Assumption 6.12*

Let  $\bar{r}$ ,  $\mu$ , and  $\mathfrak{m}_\emptyset$  be as above. Then:

- (1) We have  $\bar{r}_{\mathfrak{m}_\emptyset} \cong \bar{r}$ . In particular, since  $\bar{r}$  is absolutely irreducible,  $\mathfrak{m}_\emptyset$  is non-Eisenstein.
- (2) For each  $x \in Q$ ,  $x \equiv 1 \pmod{p}$  and  $\bar{r} \upharpoonright G_x$  is a direct sum of four pairwise distinct characters with Frobenius eigenvalues  $\alpha_x, \beta_x, \gamma_x, \delta_x$ . We assume that the eigenvalues have been labeled so that the plane  $\lambda(\alpha_x) \oplus \lambda(\beta_x)$  is isotropic, and hence  $\alpha_x \delta_x = \beta_x \gamma_x$ .

We let  $\mathfrak{m} \subset \mathbf{T}_\mu(Q)^{\text{ord}}$  be any maximal ideal which contains  $\mathfrak{m}_\emptyset$ . The representations  $\bar{r}_\mathfrak{m}$ ,  $\bar{r}_{\mathfrak{m}_\emptyset}$ , and  $\bar{r}$  are all isomorphic.

We now turn to the prime  $p$ . Let  $\alpha, \beta \in k^\times$  be the elements associated to  $\bar{r} \upharpoonright G_{\mathbf{Q}_p}$  at the beginning of Section 4. For  $M = \mathcal{O}, \mathcal{O}/\varpi^m$  or  $K/\mathcal{O}$ , we define

- $H^0(X_1(Q), \omega(a, b)(-\infty)_M)^\beta$  to be the subspace of  $H^0(X_1(Q), \omega(a, b)(-\infty)_M)$  given by the image of the idempotent  $e_\beta = \varinjlim_n ((T_{p,1} - \tilde{\beta})(Q_{p,2} - \tilde{\alpha}\tilde{\beta}))^{n!}$ , where  $\tilde{\alpha}$  and  $\tilde{\beta}$  are lifts of  $\alpha$  and  $\beta$  to  $\mathcal{O}$ ;
- $\mathbf{T}_\mu^{\text{an}}(Q)^\beta$  (resp.,  $\mathbf{T}_\mu(Q)^\beta$ ,  $\tilde{\mathbf{T}}_\mu(Q)^\beta$ ) to be the image of  $\mathbf{T}_\mu^{\text{an}}(Q)$  (resp.,  $\mathbf{T}_\mu(Q)$ ,  $\tilde{\mathbf{T}}_\mu(Q)$ ) in

$$\text{End}_\emptyset(H^0(X_1(Q), \omega(a, b)(-\infty)_M)^\beta).$$

We also make the analogous definitions with  $\alpha$  and  $\beta$  swapping roles.

**THEOREM 6.13**

Let  $\mu = (a, b)$ ,  $\mathfrak{m}_\emptyset$ , and  $\mathfrak{m}$  be as above, and suppose that Assumption 6.12 holds. Let  $w = a + b - 6$  and  $\mathbf{w} = w + 3 = a + b - 3$ . Then there exists a continuous representation

$$r = r_{\mu, \mathfrak{m}}^\beta : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathbf{T}_\mu(Q)_{\mathfrak{m}}^\beta)$$

lifting  $\bar{r}_{\mathfrak{m}} = \bar{r}$  and such that:

- (1) The similitude character  $\nu \circ r$  is given by

$$\nu \circ r = \chi_\mu \epsilon^{-3} = \chi_{\mu, 0} \epsilon^{-w},$$

where  $\chi_{\mu, 0}$  is a finite-order character unramified at  $p$  which is trivial modulo  $\mathfrak{m}$ .

- (2) For each prime  $x \notin S \cup Q \cup \{p\}$ ,  $r$  is unramified at  $x$  and  $r(\mathrm{Frob}_x)$  has characteristic polynomial  $Q_x(X)$ .
- (3) There are units  $d_{p,1}, \dots, d_{p,4} \in \mathbf{T}_\mu(Q)_{\mathfrak{m}}^\beta$  satisfying

$$\begin{aligned} Q_p(X) &= (X - d_{p,1})(X - p^{b-2}d_{p,2})(X - p^{a-1}d_{p,3})(X - p^w d_{p,4}) \\ &\in \mathbf{T}_\mu(Q)_{\mathfrak{m}}^\beta[X], \end{aligned}$$

and such that

- (a) we have  $d_{p,1} \pmod{\mathfrak{m}} = \beta$  and  $d_{p,2} \pmod{\mathfrak{m}} = \alpha$ ;
- (b)  $r \mid G_{\mathbf{Q}_p}$  is conjugate in  $\mathrm{GSp}_4$  to a representation of the form:

$$\begin{pmatrix} \lambda(d_{p,1}) & * & * & * \\ 0 & \epsilon^{-(b-2)} \cdot \lambda(d_{p,2}) & * & * \\ 0 & 0 & \epsilon^{-(a-1)} \cdot \lambda(d_{p,3}) & * \\ 0 & 0 & 0 & \epsilon^{-w} \cdot \lambda(d_{p,4}) \end{pmatrix}.$$

- (4) After twisting by the unique square-root of  $\chi_{\mu, 0}$  which is trivial modulo  $\mathfrak{m}$ , the deformation  $r$  of  $\bar{r}$  satisfies properties (2)–(5) of Definition 4.6.

*Remark 6.14*

We expect that, under the given assumptions, the Hecke rings in question are torsion-free. However, we avoid having to prove this by passing to sufficiently high weight.

*Proof*

As in Remark 6.5,  $I_{\mu, m}$  denotes the annihilator of  $H^0(X_1(Q), \omega(a, b)(-\infty)_{\mathcal{O}/\varpi^m})$  in  $\mathbf{T}_\mu(Q)$ . Since  $\mathbf{T}_\mu(Q)_{\mathfrak{m}} = \varprojlim_m \mathbf{T}_\mu(Q)_{\mathfrak{m}}/I_{\mu, m}$ , it suffices to construct, for each  $m > 0$ , a representation  $r_m : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathbf{T}_\mu(Q)_{\mathfrak{m}}^\beta/I_{\mu, m})$  satisfying the conditions of the theorem. We thus fix an  $m > 0$ . Choose an integer  $s > 0$  as in Lemma 6.4, and let  $t = rs(p - 1)$ . By Lemma 6.4 and Lemma 6.2(2), multiplication by  $A^s$  restricts to a map:

$$H^0(X_1(Q), \omega(a, b)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^\beta \hookrightarrow H^0(X_1(Q), \omega(a + t, b + t)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^\beta.$$

This in turns gives rise to a surjective map  $\mathbf{T}_{\mu'}(Q)_m^\beta \rightarrow \mathbf{T}_\mu(Q)_m^\beta / I_{\mu,m}$ . Thus it suffices to prove the result in weight  $\mu' := (a', b') := (a + t, b + t)$ .

Since  $t \geq N(\mu)$ , we have that

$$H^0(X_1(Q), \omega(a', b')(-\infty)_{K/\mathcal{O}}) \cong H^0(X_1(Q), \omega(a', b')(-\infty)) \otimes K/\mathcal{O}$$

and hence we may regard  $\mathbf{T}_{\mu'}(Q)$  as acting faithfully on both

$$H^0(X_1(Q), \omega(a', b')(-\infty)) \quad \text{and} \quad H^0(X_1(Q), \omega(a', b')(-\infty)_K).$$

Thus we have

$$\mathbf{T}_{\mu'}(Q)_m^\beta \hookrightarrow \prod_i \mathcal{O}_{K_i},$$

where the  $K_i$ 's are a finite collection of finite extensions of  $K$ , one for each minimal prime  $\wp_i$  of  $\mathbf{T}_{\mu'}(Q)_m^\beta$ . Each such minimal prime corresponds to an eigenform  $f_i$  for  $\mathbf{T}_{\mu'}(Q)_m^\beta$ . The eigenform  $f_i$  has an associated Galois representation  $r_{f_i} : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathcal{O}_{K'_i})$  for some finite extension  $K'_i/K_i$ , by Proposition 6.8. After conjugation, we may assume that each  $r_{f_i}$  reduces to  $\bar{r}$ . By the argument of the proof of [20, Proposition 3.4.4], using [29, Lemma 7.1.1] in place of [20, Lemma 2.1.12], we see that the representation  $\prod_i r_{f_i}$  descends to a representation  $r : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathbf{T}_{\mu'}(Q)_m^\beta)$ . It follows from Proposition 6.8 that  $r$  satisfies properties (1)–(3) of the theorem. For part (3), note that  $Q_p(X) \in \mathbf{T}_{\mu'}(Q)_m^\beta$  factors as

$$(X - d_{p,1})(X - p^{b-2}d_{p,2})(X - p^{a-1}d_{p,1})(X - p^w d_{p,4})$$

for units  $d_{p,i} \in \mathbf{T}_{\mu'}(Q)_m^\beta$ . We also have  $T_{p,1} \equiv \beta \pmod{m}$  and  $Q_{p,2} \equiv \alpha\beta \pmod{m}$  in  $\mathbf{T}_{\mu'}(Q)_m^\beta$  (by definition of the idempotent  $e_\beta$ ). Since  $Q_p(X) = X^4 - T_{p,1}X^3 + p^{b-2}Q_{p,2}X^2 - \dots$ , we deduce that  $d_{p,1} \equiv \beta \pmod{m}$  and  $d_{p,2} \equiv \alpha \pmod{m}$ .

To show that  $r$  satisfies properties (2)–(5) of Definition 4.6, it suffices to show that each  $r_{f_i}$  does so. In fact, property (2) has already been established with the exception of the prime  $x = q$ . If  $x = q$ , then (by our assumptions)  $\mathrm{ad}^0(\bar{r})(1)$  as a  $G_{\mathbf{Q}_q}$ -module contains no subquotient isomorphic to  $k$ , and so  $H^2(\mathbf{Q}_q, \mathrm{ad}^0(\bar{r})) \simeq H^0(\mathbf{Q}_q, \mathrm{ad}^0(\bar{r})(1))^* = 0$ . Since  $q \neq p$ , it follows that  $H^1(\mathbf{Q}_q, \mathrm{ad}^0(\bar{r}))$  consists entirely of unramified classes. In particular, all lifts of  $\bar{r}$  are automatically unramified at  $q$ . Since  $m$  is non-Eisenstein, it follows from Proposition 6.8(5) that  $r_{f_i}$  satisfies local-global compatibility at all primes. Thus we may apply the results of [64, Section 4.5]. We now turn to property (3) of Definition 4.6. If  $x \in S(\bar{r})$  is of type **U3**, then  $\bar{r}(I_x)$  is unipotent and generated by a conjugate of  $\exp(N_3)$ . Since  $K_x = I(x)$ , [64, Corollary 1] implies that  $r_{f_i}(I_x)$  is topologically generated by a conjugate of  $\exp(N_3)$ ,  $\exp(N_2)$ , or  $\exp(N_1)$ . The latter two cases are incompatible

with the residual representation being of nilpotent rank 3. Similarly, if  $x \in S(\bar{F})$  is of type **U2**, then  $K_x = \Pi(x)$  and [64, Corollary 1] implies that  $r_{f_i}(I_x)$  is topologically generated by a conjugate of  $\exp(N_2)$  or  $\exp(N_1)$ . The latter case is incompatible with the residual representation being of nilpotent rank 2. Finally, if  $x \in S(\bar{F})$  is of type **U1**, then  $K_x = K(x)$ . It then suffices to note, following [64, Section 4.5], that the corresponding representation  $\pi_x$  is *paraspherical*, that is, has a nonzero fixed vector by a nonspecial maximal compact subgroup, namely,  $K(x)$  itself. This establishes property (3). For property (4), suppose that  $x \in S(\bar{F})$  is of type **P**. Then  $K_x = \Pi(x)^+$ . It follows from [64, Corollary 1] that  $\Pi(x)$  has no invariants on the automorphic representation generated by  $f_i$  (as otherwise  $r_{f_i} | I_x$  would be unipotent, contradicting the assumption on  $\bar{F}$  at  $x$ ). Thus  $\Pi(x)/\Pi(x)^+$  acts through a nontrivial character on the space of  $\Pi(x)^+$  invariants. By [64, Corollary 3] all such characters have to lift the character  $\nu \circ \bar{F} | I_x$ . However, since  $x - 1$  is prime to  $p$ , there is a unique such character, and the result follows from [64, Corollary 3].

Finally, we turn to Definition 4.6(5). Let  $x \in Q$ , and recall that  $K_x = \Pi(x)^+$ . Let  $\pi$  be the automorphic representation generated by  $f_i$ . Consider first the case where  $\pi_x$  has nontrivial  $\Pi(x)$ -invariants. Then  $\pi_x$  is a subquotient of an unramified principal series. By Assumption 6.12(2) and [31, Proposition 3.2.3], we see that  $\pi_x$  is unramified. In this case, Definition 4.6(5) certainly holds for  $r_{f_i}$ . In the remaining case, where  $\pi_x$  has no nontrivial  $\Pi(x)$ -invariants, we see that  $\Pi(x)/\Pi^+(x)$  acts through a nontrivial character on  $\pi_x^{\Pi(x)^+}$ , and the required property holds by [64, Corollary 3].  $\square$

#### 6.4. Galois representations in low weights

We let  $\bar{r}: G_{\mathbb{Q}} \rightarrow \mathrm{GSp}_4(k)$ ,  $S = S(\bar{r})$ ,  $Q$ , and  $K \subset \mathrm{GSp}_4(\mathbb{A}^{\infty})$  be as in the previous section. Recall that in Section 4, we fixed two units  $\alpha, \beta \in k^{\times}$  associated to  $\bar{r} | G_{\mathbb{Q}, p}$ . We now let  $\sigma = (a, 2) \in X^*(T)_M^+$  with  $a \geq 2$  denote a nonregular weight.

##### Definition 6.15

We say that  $\bar{r}$  is *Katz modular of weight  $\sigma$*  if there exists a maximal ideal  $\mathfrak{m}_{\theta}$  of  $\mathbf{T}_{\sigma}$  such that

- (1) we have  $\bar{r}_{\mathfrak{m}_{\theta}} \cong \bar{r}$ , and
- (2) there exists a form  $\eta \in H^0(X, \omega(a, 2)_{K/\theta})[\mathfrak{m}_{\theta}]$  such that

$$T_{p,1}(\eta) = (\alpha + \beta)\eta,$$

$$Q_{p,2}(\eta) = (\alpha\beta)\eta.$$

We now make the following assumption.

*Assumption 6.16 (Residual modularity)*

We assume the following.

- (1) The representation  $\bar{r}$  is Katz modular of weight  $\sigma$  with associated maximal ideal  $\mathfrak{m}_\emptyset$  and eigenform  $\eta$ .
- (2) For each  $x \in Q$ ,  $x \equiv 1 \pmod{p}$  and  $\bar{r} \upharpoonright G_x$  is a direct sum of four pairwise distinct characters with Frobenius eigenvalues  $\alpha_x, \beta_x, \gamma_x, \delta_x$ . We assume that the eigenvalues have been labeled so that the plane  $\lambda(\alpha_x) \oplus \lambda(\beta_x)$  is isotropic, and hence  $\alpha_x \delta_x = \beta_x \gamma_x$ .

We let  $\mathfrak{m}$  be any maximal ideal of  $\mathbf{T}_\sigma(Q)$  containing  $\mathfrak{m}_\emptyset$ .

Let  $e_{\alpha,\beta}$  be the idempotent

$$\lim_{\substack{\longrightarrow \\ n}} ((T_{p,1} - \tilde{\alpha} - \tilde{\beta})(Q_{p,2} - \tilde{\alpha}\tilde{\beta}))^{n!},$$

where  $\tilde{\alpha}$  and  $\tilde{\beta}$  are lifts of  $\alpha$  and  $\beta$  to  $\mathcal{O}$ , and define

$$H^0(X_1(Q), \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha,\beta} = e_{\alpha,\beta} H^0(X_1(Q), \omega(a, 2)(-\infty)_{K/\mathcal{O}}).$$

The assumption that  $\bar{r}$  is Katz modular implies that this space is nonzero after localization at  $\mathfrak{m}$ . We let  $\mathbf{T}_\sigma(Q)^{\alpha,\beta}$  denote the image of  $\mathbf{T}_\sigma(Q)$  in

$$\text{End}_{\mathcal{O}}(H^0(X_1(Q), \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha,\beta}).$$

Our main result in this section is the following.

**THEOREM 6.17**

Let  $\bar{r}, \sigma = (a, 2)$  with  $p-1 > a$  and  $\mathfrak{m}$  be as above, and suppose that Assumption 6.16 holds. In addition, suppose that

$$(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2 \beta^2 - 1) \neq 0.$$

Then there exists a representation

$$r_Q : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\mathbf{T}_\sigma(Q)_{\mathfrak{m}}^{\alpha,\beta})$$

which is a minimal deformation of  $\bar{r}$  outside  $Q$ .

*Proof*

As in the proof of Theorem 6.13, it suffices to prove the existence of an appropriate representation  $r_m : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\mathbf{T}_\sigma(Q)_{\mathfrak{m}}^{\alpha,\beta} / I_{\sigma,m})$  for each  $m > 0$ . We thus fix an  $m > 0$ . By Theorem 8.13 below, there exists a power  $A^s$  of  $A$  such that we have injections

$$H^0(X_1(Q), \omega(a, 2)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\alpha, \beta} \xrightarrow{e_{\beta \circ A^s}} H^0(X_1(Q), \omega(a+t, b+t)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\beta},$$

$$H^0(X_1(Q), \omega(a, 2)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\alpha, \beta} \xrightarrow{e_{\alpha \circ A^s}} H^0(X_1(Q), \omega(a+t, b+t)(-\infty)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\alpha},$$

where  $t = rp^{m-1}(p-1)$ . These in turns give rise to surjections

$$\mathbf{T}_{\mu'}(Q)_{\mathfrak{m}}^{\beta} \twoheadrightarrow \mathbf{T}_{\mu}(Q)_{\mathfrak{m}}^{\alpha, \beta} / I_{\mu, m},$$

$$\mathbf{T}_{\mu'}(Q)_{\mathfrak{m}}^{\alpha} \twoheadrightarrow \mathbf{T}_{\mu}(Q)_{\mathfrak{m}}^{\alpha, \beta} / I_{\mu, m},$$

where  $\mu' = \mu + (t, t)$ . The first of these surjections together with Theorem 6.13 implies the existence of a representation  $r'_m$  satisfying all of the required properties, except for conditions (1) and (6) of Definition 4.6. However, we deduce from the existence of both surjections that the representation  $r_m | G_p$  contains two distinct rank 1 unramified submodules (spanned by basis vectors), with one having Frobenius eigenvalue lifting  $\alpha$  and the other having Frobenius eigenvalue lifting  $\beta$ . By Nakayama's lemma, we deduce that  $r'_m$  contains an unramified rank 2 submodule of the form required by Definition 4.6(6). In order to obtain a representation that also satisfies Definition 4.6(1), we note that  $\nu(r'_m) = \chi \epsilon^{-(a-1)} \chi_Q$ , where  $\chi_Q$  is a finite-order character of  $p$ -power order which is unramified outside  $Q$ . Since  $p$  is odd, we can find a square root of  $\chi_Q$  and twist  $r'_m$  by the inverse of this square root. The resulting representation  $r_m$  now satisfies all required properties.  $\square$

## 7. Properties of cohomology groups

As in Section 5.2, let  $S$  and  $Q$  be finite sets of primes of  $\mathbf{Q}$  which are disjoint and do not contain  $p$ . We allow the possibility that  $Q = \emptyset$ . We let  $K$  and  $K_i(Q)$  be open compact subgroups of  $\mathrm{GSp}_4(\mathbb{A}^{\infty})$  as in Section 5.2, and we let  $X = X_K$  and  $X_i(Q) = X_{K_i(Q)}$  be the corresponding Siegel 3-folds. The goal of this section is to prove Theorems 7.2 and 7.11 below.

### 7.1. Taylor–Wiles primes

Fix  $\mu = (a, b) \in X^*(T)_M^+$  with  $a \geq b \geq 2$ . Let  $\mathfrak{m}_{\emptyset}$  be a non-Eisenstein maximal ideal of  $\mathbf{T}_{\mu}$ . The ideal  $\mathfrak{m}_{\emptyset}$  gives rise to ideals of  $\mathbf{T}_{\mu}^{\mathrm{an}}(Q)$  and  $\mathbf{T}_{\mu}(Q)$  which we also denote by  $\mathfrak{m}_{\emptyset}$  (see Section 6.3). We will need the following assumption (cf. Assumptions 6.12, 6.16).

#### *Assumption 7.1*

For each  $x \in Q$ , we have  $x \equiv 1 \pmod{p}$ , and  $\bar{r}_{\mathfrak{m}_{\emptyset}} | G_x$  is a direct sum of four pairwise distinct characters with Frobenius eigenvalues  $\alpha_x, \beta_x, \gamma_x, \delta_x$ . We assume that the eigenvalues have been labeled so that the plane  $\lambda(\alpha_x) \oplus \lambda(\beta_x)$  is isotropic, and hence  $\alpha_x \delta_x = \beta_x \gamma_x$ .

For  $x \in Q$ , we let  $\alpha'_x, \beta'_x, \gamma'_x, \delta'_x \in \mathcal{O}^\times$  be elements lifting  $\alpha_x, \beta_x, \gamma_x, \delta_x \in k^\times$ . The point of the above assumption is to rule out the possibility of newforms at level  $K_0(Q)$ .

**THEOREM 7.2**

Let  $\mu$  and  $\mathfrak{m}_\emptyset$  be as above, and suppose that Assumption 7.1 holds. Let  $\mathfrak{m}$  denote the ideal of  $\mathbf{T}_\mu(Q)$  containing  $\mathfrak{m}_\emptyset$  together with the elements  $xU_{x,2} - \alpha'_x\beta'_x$  and  $U_{x,1} - \alpha'_x - \beta'_x$  for each  $x \in Q$ . Then  $\mathfrak{m}$  is maximal and there is an isomorphism

$$\mathrm{pr}_Q \circ i : H^0(X, \omega(a, b)(-\infty)_{K/\mathcal{O}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_0(Q), \omega(a, b)(-\infty)_{K/\mathcal{O}})_{\mathfrak{m}}$$

which is equivariant for the operators  $T_{x,i}$  for each  $x \notin S \cup Q \cup \{p\}$  as well as for the operators  $T_{p,1}$  and  $Q_{p,2}$ .

Here  $i$  is the natural inclusion and  $\mathrm{pr}_Q$  is defined as follows. For  $x \in Q$ , let  $R_x$  denote the Hecke operator

$$R_x = (xU_{x,2} - \alpha'_x\gamma'_x)(xU_{x,2} - \beta'_x\delta'_x)(xU_{x,2} - \gamma'_x\delta'_x) \in \mathbf{T}_\mu(Q),$$

and let  $\mathrm{pr}_x$  denote the idempotent

$$\mathrm{pr}_x = \lim_{n \rightarrow \infty} (R'_x)^{n!}.$$

Then the  $\mathrm{pr}_x$ 's commute with one another and  $\mathrm{pr}_Q$  denotes their product.

For compactness, we will make use of the alternative notation  $\mathcal{W}_\mu = \omega(a, b)$  and  $\mathcal{W}_\mu^{\mathrm{sub}} = \omega(a, b)(-\infty)$ . In sufficiently high weight, Theorem 7.2 is due to Genestier and Tilouine.

**THEOREM 7.3**

Suppose that  $\mu = (a, b)$  is such that  $H^i(X, \mathcal{W}_{\mu,k}^{\mathrm{sub}})$  and  $H^i(X_0(Q), \mathcal{W}_{\mu,k}^{\mathrm{sub}})$  are zero for all  $i > 0$ . Then the map

$$\mathrm{pr}_Q \circ i : H^0(X, \mathcal{W}_{\mu,K/\mathcal{O}}^{\mathrm{sub}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(X_0(Q), \mathcal{W}_{\mu,K/\mathcal{O}}^{\mathrm{sub}})_{\mathfrak{m}}$$

is an isomorphism. An explicit inverse is given by the composition

$$H^0(X_0(Q), \mathcal{W}_{\mu,K/\mathcal{O}}^{\mathrm{sub}})_{\mathfrak{m}} \hookrightarrow H^0(X_0(Q), \mathcal{W}_{\mu,K/\mathcal{O}}^{\mathrm{sub}})_{\mathfrak{m}_\emptyset} \xrightarrow{d_Q^{-1} \mathrm{tr}} H^0(X, \mathcal{W}_{\mu,K/\mathcal{O}}^{\mathrm{sub}})_{\mathfrak{m}_\emptyset},$$

where  $d_Q = \prod_{x \in Q} [\mathrm{GSp}_4(\mathbf{Z}_x) : \Pi(x)]$  (which is prime to  $p$ ) and  $\mathrm{tr}$  is the trace map associated to  $X_0(Q) \rightarrow X$ .

*Proof*

By the assumption of cohomology vanishing, it suffices to prove both statements with  $K/\mathcal{O}$  replaced by  $K$ . Indeed, if the map over  $K$  is surjective, then so too is the

map over  $K/\mathcal{O}$ . Furthermore, if  $d_Q^{-1}\text{tr}$  is an inverse over  $K$ , then the fact that it is defined over  $\mathcal{O}$  implies immediately that it also gives an inverse over  $K/\mathcal{O}$ . The proof of the corresponding result over  $K$  follows exactly as in the proof of [31, Proposition 11.1.2].  $\square$

Using this result and the Hasse invariant  $h \in H^0(X, \omega_k^{p-1})$ , we can now establish Theorem 7.2 at the level of  $\varpi$ -torsion. (Recall that cohomology in degree 0 over  $k$  can be identified with  $\varpi$ -torsion in degree 0 cohomology over  $K/\mathcal{O}$ .) Note that, for any weight  $\mu$ , the cohomology vanishing assumption of the previous theorem holds in weight  $\mu + (t, t)$  as long as  $t \geq N(\mu)$  (where  $N(\mu)$  is defined in Definition 6.3).

LEMMA 7.4

Let  $\mu = (a, b)$  with  $a \geq b \geq 2$ . Choose an integer  $t$  such that  $(p-1)t \geq N(\mu)$ . Let  $\mu' = (a', b') = (a + t(p-1), b + t(p-1))$ . Then the following diagrams are co-Cartesian:

$$\begin{array}{ccc}
 H^0(X_0(Q), \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}} & \xrightarrow{h^t} & H^0(X_0(Q), \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}} \\
 \text{pr}_Q \circ i \uparrow & & \text{pr}_Q \circ i \uparrow \cong \\
 H^0(X, \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} & \xrightarrow{h^t} & H^0(X, \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} \\
 \\ 
 H^0(X_0(Q), \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}} & \xrightarrow{h^t} & H^0(X_0(Q), \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}} \\
 d_Q^{-1}\text{tr} \downarrow & & d_Q^{-1}\text{tr} \downarrow \cong \\
 H^0(X, \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} & \xrightarrow{h^t} & H^0(X, \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}}
 \end{array}$$

In particular, the left-hand vertical maps are mutually inverse isomorphisms.

*Proof*

Note that the right-hand vertical maps are mutually inverse isomorphisms by Theorem 7.3 and the choice of  $t$ . The diagrams are commutative because  $h$  commutes with all Hecke operators at the primes in  $Q$  (see Lemma 6.2). Now, let  $f \in H^0(X, \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}}$ , and let  $F = \text{pr}_Q(f) \in H^0(X_0(Q), \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}}$ . Note that  $f$  can be recovered from  $F$  via the formula  $f = d_Q^{-1}\text{tr}(F)$ . We need to show that  $f$  is divisible by  $h^t$  if and only if  $F$  is divisible by  $h^t$ . But this follows immediately by the commutativity of the diagrams above: if  $f = h^t g$ , then  $F = h^t \text{pr}_Q(g)$ , and if



$F = h^t g$ , then  $f = h^t d_Q^{-1} \text{tr}(g)$ . (Note that since  $X_0(Q)$  and  $X$  are smooth (and in particular irreducible) over  $k$ , multiplication by  $h$  is injective on  $H^0$ .)  $\square$

We will need the analogous result for forms on the nonordinary locus. Let  $S$  (resp.,  $S_0(Q)$ ) denote the nonordinary locus of  $X_k$  (resp.,  $X_0(Q)_k$ ).

LEMMA 7.5

Let  $\mu = (a, b)$  with  $a \geq b \geq 2$ . Then the map

$$\text{pr}_Q \circ i : H^0(S, \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}_\emptyset} \xrightarrow{\sim} H^0(S_0(Q), \mathcal{W}_{\mu, k}^{\text{sub}})_{\mathfrak{m}}$$

is an isomorphism with inverse  $d_Q^{-1} \text{tr}$ .

*Proof*

We first show that the result is true in sufficiently high weight. More precisely: let  $t \geq N(\mu) + 1$ . We let  $\mu' = (a + (t - 1)(p - 1), b + (t - 1)(p - 1))$  and  $\mu'' = (a + t(p - 1), b + t(p - 1))$ . We have a commutative diagram:

$$\begin{array}{ccccccc} H^0(X_0(Q), \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}} & \xrightarrow{h} & H^0(X_0(Q), \mathcal{W}_{\mu'', k}^{\text{sub}})_{\mathfrak{m}} & \longrightarrow & H^0(S_0(Q), \mathcal{W}_{\mu'', k}^{\text{sub}})_{\mathfrak{m}} & \longrightarrow & 0 \\ \uparrow \text{pr}_Q \cong & & \uparrow \text{pr}_Q \cong & & \uparrow \text{pr}_Q & & \\ H^0(X, \mathcal{W}_{\mu', k}^{\text{sub}})_{\mathfrak{m}_\emptyset} & \xrightarrow{h} & H^0(X, \mathcal{W}_{\mu'', k}^{\text{sub}})_{\mathfrak{m}_\emptyset} & \longrightarrow & H^0(S, \mathcal{W}_{\mu'', k}^{\text{sub}})_{\mathfrak{m}_\emptyset} & \longrightarrow & 0 \end{array}$$

The choice of  $t$  guarantees that the rows are short exact sequences. From the previous lemma, we deduce that the right-hand vertical map is an isomorphism with inverse  $d_Q^{-1} \text{tr}$ .

Now we imitate the proof of the previous lemma to deduce the result in smaller weights. For this we use the existence of the Hasse invariant

$$\tilde{h} \in H^0(S, \omega(p^2 - 1, p^2 - 1)_k).$$

Such a form was constructed in unpublished work of the second author with Goldring, but is also constructed in greater generality in [12] and [33]. In [12, Theorem B.2] (see also [12, Theorem 6.2.3]), it is shown that  $\tilde{h}$  extends to the boundary (by the normality of the  $p$ -rank 1 locus) and that multiplication by  $\tilde{h}$  is Hecke equivariant away from  $p$  (see [12, Theorem 4.5.4(3)]). (It is also true, but not relevant here, that  $\tilde{h}$  vanishes on the 1-dimensional Ekedahl–Oort stratum of  $S$  to precise order 2; see the references in the proof of Theorem 8.10 below for more discussion on this point.)

We choose an integer  $s$  such that  $t := s(p + 1) \geq N(\mu) + 1$ . Let  $\mu'' = \mu + s(p^2 - 1, p^2 - 1) = \mu + t(p - 1, p - 1)$ . Then we have a commutative diagram:

$$\begin{array}{ccc}
H^0(S_0(Q), \mathcal{W}_{\mu,k}^{\text{sub}})_m & \xrightarrow{\tilde{h}^s} & H^0(S_0(Q), \mathcal{W}_{\mu'',k}^{\text{sub}})_m \\
\text{pr}_Q \circ i \uparrow & & \text{pr}_Q \circ i \uparrow \cong \\
H^0(S, \mathcal{W}_{\mu,k}^{\text{sub}})_{m\emptyset} & \xrightarrow{\tilde{h}^s} & H^0(S, \mathcal{W}_{\mu'',k}^{\text{sub}})_{m\emptyset}
\end{array}$$

The right-hand vertical map is an isomorphism with inverse  $d_Q^{-1}\text{tr}$  by the first paragraph. The lemma now follows by the same argument as Lemma 7.4.  $\square$

We will also need the following analogous result for first degree cohomology over  $k$ .

LEMMA 7.6

Suppose that  $\mu = (a, b)$  where  $a \geq b \geq 2$ . Then the map

$$\text{pr}_Q \circ i : H^1(X, \mathcal{W}_{\mu,k}^{\text{sub}})_{m\emptyset} \xrightarrow{\sim} H^1(X_0(Q), \mathcal{W}_{\mu,k}^{\text{sub}})_m$$

is an isomorphism with inverse  $d_Q^{-1}\text{tr}$ .

*Proof*

If  $N(\mu) = 0$ , then both sides of the map are zero, so we may assume that  $N(\mu) > 0$ . Let  $t \geq N(\mu)$ , and let  $\mu' = (a + (t-1)(p-1), b + (t-1)(p-1))$  and  $\mu'' = (a + t(p-1), b + t(p-1))$ . Consider the diagram with exact rows:

$$\begin{array}{ccccccc}
H^0(X_0(Q), \mathcal{W}_{\mu',k}^{\text{sub}})_m & \xrightarrow{h} & H^0(X_0(Q), \mathcal{W}_{\mu'',k}^{\text{sub}})_m & \longrightarrow & H^0(S_0(Q), \mathcal{W}_{\mu'',k}^{\text{sub}})_m & \longrightarrow & H^1(X_0(Q), \mathcal{W}_{\mu',k}^{\text{sub}})_m \longrightarrow 0 \\
\text{pr}_Q \uparrow \cong & & \text{pr}_Q \uparrow \cong & & \text{pr}_Q \uparrow \cong & & \text{pr}_Q \uparrow \cong \\
H^0(X, \mathcal{W}_{\mu',k}^{\text{sub}})_{m\emptyset} & \xrightarrow{h} & H^0(X, \mathcal{W}_{\mu'',k}^{\text{sub}})_{m\emptyset} & \longrightarrow & H^0(S, \mathcal{W}_{\mu'',k}^{\text{sub}})_{m\emptyset} & \longrightarrow & H^1(X, \mathcal{W}_{\mu',k}^{\text{sub}})_{m\emptyset} \longrightarrow 0
\end{array}$$

The first three vertical maps are isomorphisms with inverse  $d_Q^{-1}\text{tr}$  by Lemmas 7.4 and 7.5. We deduce that the rightmost vertical map above is an isomorphism with inverse  $d_Q^{-1}\text{tr}$ . This proves the lemma in weight  $\mu'$ . The general case then follows by a similar argument using a reverse induction on  $t$ .  $\square$

We are finally in a position to prove Theorem 7.2 in the general case.

*Proof of Theorem 7.2*

For each  $n \geq 1$ , let  $\mathcal{O}_n := \mathcal{O}/\varpi^n$ . We have a commutative diagram:

$$\begin{array}{ccccccc}
 H^0(X_0(Q), \mathcal{W}_{\mu,k}^{\text{sub}})_{\mathfrak{m}} & \longrightarrow & H^0(X_0(Q), \mathcal{W}_{\mu,\mathcal{O}_n}^{\text{sub}})_{\mathfrak{m}} & \xrightarrow{\varpi} & H^0(X_0(Q), \mathcal{W}_{\mu,\mathcal{O}_{n-1}}^{\text{sub}})_{\mathfrak{m}} & \longrightarrow & H^1(X_0(Q), \mathcal{W}_{\mu,k}^{\text{sub}})_{\mathfrak{m}} \\
 \uparrow \text{pr}_{\mathcal{Q}} \cong & & \uparrow \text{pr}_{\mathcal{Q}} & & \uparrow \text{pr}_{\mathcal{Q}} & & \uparrow \text{pr}_{\mathcal{Q}} \cong \\
 H^0(X, \mathcal{W}_{\mu,k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} & \longrightarrow & H^0(X, \mathcal{W}_{\mu,\mathcal{O}_n}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} & \xrightarrow{\varpi} & H^0(X, \mathcal{W}_{\mu,\mathcal{O}_{n-1}}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} & \longrightarrow & H^1(X, \mathcal{W}_{\mu,k}^{\text{sub}})_{\mathfrak{m}_{\emptyset}}
 \end{array}$$

The vertical maps on the ends are isomorphisms by Lemmas 7.4 and 7.6. By induction on  $n$  and the five lemma, we deduce that the map

$$\text{pr}_{\mathcal{Q}} \circ i : H^0(X, \mathcal{W}_{\mu,\mathcal{O}_n}^{\text{sub}})_{\mathfrak{m}_{\emptyset}} \rightarrow H^0(X_0(Q), \mathcal{W}_{\mu,k}^{\text{sub}})_{\mathfrak{m}}$$

is an isomorphism for all  $n$ . This shows that the map of Theorem 7.2 is an isomorphism after passing to  $\varpi^n$ -torsion, for any  $n$ . The result follows.  $\square$

### 7.2. The balanced property

In this section we assume that  $\mu = (a, 2)$  is a limit of discrete series weight, where  $p > a - 2$ . Let  $\Delta$  be a quotient of  $\Delta_{\mathcal{Q}} := \prod_{x \in \mathcal{Q}} (\mathbf{Z}/x)^{\times}$ , and let  $X_{\Delta}(Q) \rightarrow X_0(Q)$  denote the corresponding subcover of  $X_1(Q) \rightarrow X_0(Q)$ . If  $\mathcal{L}$  is a vector bundle on  $X_{\Delta}(Q)$ , then we define

$$H_i(X_{\Delta}(Q), \mathcal{L}) := H^i(X_{\Delta}(Q), (\omega^3 \otimes \mathcal{L}^{\vee}(-\infty))_{K/\mathcal{O}})^{\vee}$$

for all  $i$ . Note that  $\omega^3(-\infty)$  is the dualizing sheaf on  $X_{\Delta}(Q)$ .

We now take  $\mathcal{L} = \omega(1, 3 - a)$ , so that  $\omega^3 \otimes \mathcal{L}^{\vee}(-\infty) \cong \omega(a, 2)(-\infty)$ . Here we use our bound  $p > a - 2$  to deduce that there is an equality  $(\text{Sym}^{a-2})^{\vee} \simeq \text{Sym}^{a-2} \otimes \det^{2-a}$  as  $\mathcal{O}$ -modules. Thus,  $\mathbf{T}_{\mu}(Q)$  acts on  $H_0(X_{\Delta}(Q), \omega(1, 3 - a))$ . We fix a non-Eisenstein maximal ideal  $\mathfrak{m}$  of  $\mathbf{T}_{\mu}(Q)$ . We will need the following assumption.

#### Assumption 7.7

The space  $H^2(X_{\Delta}(Q), \omega(a, 2)(-\infty)_k)_{\mathfrak{m}}$  is trivial.

There is a slight abuse of notation here in that  $\mathbf{T}_{\mu}(Q)$  does not act on  $H^2(X_{\Delta}(Q), \omega(a, 2)_k)$ . The localization at  $\mathfrak{m}$  refers to the localization at the corresponding maximal ideal of the polynomial ring over  $\mathcal{O}$  generated by the Hecke operators.

#### Remark 7.8

We note that if  $p \geq a \geq 4$ , then the assumption above holds, even before localization at  $\mathfrak{m}$ , by Theorem 5.1.

#### LEMMA 7.9

Suppose that Assumption 7.7 holds. Then  $H_1(X_{\Delta}(Q), \omega(1, 3 - a))_{\mathfrak{m}}$  is  $p$ -torsion free.

*Proof*

The claim is equivalent to the divisibility of  $H^1(X_\Delta(Q), \omega(a, 2)(-\infty)_{K/\mathcal{O}})_\mathfrak{m}$ . Since  $X_\Delta(Q)$  is flat over  $\mathcal{O}$ , there is an exact sequence

$$0 \rightarrow \omega(a, 2)(-\infty)_k \rightarrow \omega(a, 2)(-\infty)_{K/\mathcal{O}} \xrightarrow{\varpi} \omega(a, 2)(-\infty)_{K/\mathcal{O}} \rightarrow 0.$$

Taking cohomology, this reduces to the claim that  $H^2(X_\Delta(Q), \omega(a, 2)(-\infty)_k)_\mathfrak{m}$  vanishes.  $\square$

The following lemma uses only the assumption that  $\mathfrak{m}$  is non-Eisenstein: it holds in all weights and in all prime-to- $p$  levels. We just state it in the case we need.

LEMMA 7.10

*The map*

$$H^i(X_0(Q), \omega(a, 2)(-\infty))_\mathfrak{m} \otimes K \longrightarrow H^i(X_0(Q), \omega(a, 2))_\mathfrak{m} \otimes K$$

*is an isomorphism for all  $i$ .*

*Proof*

Let  $\partial X$  denote the boundary of  $X_0(Q)$ . It suffices to show that the boundary cohomology

$$H^i(\partial X, \omega(a, 2))_\mathfrak{m} \otimes K$$

vanishes for all  $i$ . However, over  $\mathbf{C}$  the cohomology of the boundary is computed by the nerve spectral sequence (see [40, (3.2.4)])

$$E_1^{r,s} = \bigoplus_{r(R)=r+1} E_1^{r,s}(R) \implies H^{r+s}(\partial X, \omega(a, 2)_\mathbf{C}).$$

Here  $R$  is a  $\mathbf{Q}$ -parabolic of  $G$  and  $r(R)$  is its parabolic rank. By [40, Corollary 3.2.9], and freely using the notation of this article, the space  $E_1(R)^{r,s}$  is the space of  $K$ -invariants in

$$\mathrm{Ind}_{R(\mathbb{A}^\infty)}^{G(\mathbb{A}^\infty)} \bigoplus_{i \geq 0, w \in W^{R,p}} I^R(\tilde{H}^{s-i-\ell(w)}(X(G_{h,R}), \mathcal{V}_{\lambda(h,w)}) \otimes H^i(X(G_{\ell,R}), \tilde{\mathbf{V}}_{\lambda(\ell,w)})).$$

If  $R = \Pi$  is the Klingen parabolic, then  $G_{h,R} = \mathrm{GSp}_2 = \mathrm{GL}_2$  and  $G_{\ell,R} = \mathrm{GL}_1$ . If  $R$  is the Siegel parabolic or the Borel subgroup, then  $G_{h,R}$  is trivial and  $G_{\ell,R} = L_R$  is the Levi component of  $R$  (and hence is either  $\mathrm{GL}_2 \times \mathrm{GL}_1$  or  $\mathrm{GL}_1^3$ ). In all cases,  $\mathcal{V}_{\lambda(h,w)}$  is the canonical extension of an automorphic vector bundle on the Shimura variety  $X(G_h)$  and  $\tilde{\mathbf{V}}_{\lambda(\ell,w)}$  is a local system on  $X(G_\ell)$  associated to an algebraic

representation of  $G_\ell$ . See [39, (3.6.1)] for the highest weight formulas. The functor  $I_R$  is an intermediate induction defined in [40, (3.2.8)].

Since each of the groups  $G_h$  and  $G_\ell$  are products of copies of  $\mathrm{GL}_2$  and  $\mathrm{GL}_1$ , we see that to any Hecke eigenclass in any  $H^i(\partial X, \omega(a, 2)_{\mathbb{C}})$ , we can associate a compatible system of reducible  $\mathrm{GSp}_4$ -valued  $l$ -adic representations of  $G_Q$ . Since the ideal  $\mathfrak{m}$  is non-Eisenstein, it follows that  $H^i(\partial X, \omega(a, 2))_{\mathfrak{m}} \otimes K = 0$ , as required.  $\square$

We come to the main result of this section.

**THEOREM 7.11**

Let  $\Delta$  be a quotient of  $\Delta_Q$  which is of  $p$ -power order. As above, let  $\mu = (a, 2)$  with  $p-2 > a$ , and let  $\mathfrak{m}$  be a non-Eisenstein ideal of  $\mathbf{T}_\mu(Q)$ . Suppose that Assumption 7.7 holds. Then the  $\mathcal{O}[\Delta]$ -module

$$H_0(X_\Delta(Q), \omega(1, 3-a))_{\mathfrak{m}} = H^0(X_\Delta(Q), \omega(a, 2)(-\infty)_{K/\mathcal{O}})_{\mathfrak{m}}^\vee$$

is balanced in the sense of Definition 3.2.

*Proof*

The argument proceeds exactly as in the proof of [16, Proposition 3.8]. If we let  $M = H_0(X_\Delta(Q), \omega(1, 3-a))_{\mathfrak{m}}$  and  $S = \mathcal{O}[\Delta]$ , then the defect  $d_S(M)$  is given by

$$d_S(M) = r - \dim_k \mathrm{Tor}_1^S(M, \mathcal{O})/\varpi,$$

where  $r$  is the  $\mathcal{O}$ -rank of  $M_\Delta$ . Thus we need to show that  $r \geq \dim_k \mathrm{Tor}_1^S(M, \mathcal{O})/\varpi$ .

Let  $\mathcal{L} = \omega(1, 3-a)$ . Applying Pontryagin duality to the Hochschild–Serre spectral sequence, we get a spectral sequence

$$\mathrm{Tor}_i^S(H_j(X_\Delta(Q), \mathcal{L})_{\mathfrak{m}}, \mathcal{O}) \implies H_{i+j}(X_0(Q), \mathcal{L})_{\mathfrak{m}}.$$

This spectral sequence tells us that

- (1)  $M_\Delta \xrightarrow{\sim} H_0(X_0(Q), \mathcal{L})_{\mathfrak{m}}$ , and
- (2) we have a short exact sequence

$$(H_1(X_\Delta(Q), \mathcal{L})_{\mathfrak{m}})_\Delta \longrightarrow H_1(X_0(Q), \mathcal{L})_{\mathfrak{m}} \longrightarrow \mathrm{Tor}_1^S(M, \mathcal{O}) \longrightarrow 0.$$

To prove that  $d_S(M) \geq 0$ , it follows from the second point that it is sufficient to show that  $H_1(X_0(Q), \mathcal{L})_{\mathfrak{m}}$  is free of rank at most  $r$  over  $\mathcal{O}$ . Lemma 7.9 tells us that this space is  $p$ -torsion free. Passing to characteristic 0 and using the first point, we are therefore reduced to establishing the inequality

$$\dim_K H_1(X_0(Q), \mathcal{L})_{\mathfrak{m}} \otimes K \leq \dim_K H_0(X_0(Q), \mathcal{L})_{\mathfrak{m}} \otimes K.$$

In other words, we need to show that

$$\begin{aligned} & \dim_K H^1(X_0(Q), \omega(a, 2)(-\infty))_{\mathfrak{m}} \otimes K \\ & \leq \dim_K H^0(X_0(Q), \omega(a, 2)(-\infty))_{\mathfrak{m}} \otimes K. \end{aligned}$$

By Lemma 7.10, we are reduced to showing that

$$\dim_K \overline{H}^1(X_0(Q), \omega(a, 2))_{\mathfrak{m}} \otimes K \leq \dim_K \overline{H}^0(X_0(Q), \omega(a, 2))_{\mathfrak{m}} \otimes K,$$

where  $\overline{H}^i$  denotes the interior cohomology (the image of  $H^i(\omega(a, b)(-\infty))$  in  $H^i(\omega(a, b))$ ).

As recalled in Theorem 5.2, the interior cohomology can be computed in terms of square-integrable automorphic forms on  $G$ . By Remark 5.10, the cohomology of  $\omega(a, b)$  agrees with that of  $\mathcal{W}_\mu \cong \mathcal{V}_\sigma$ , where  $\mu = (a, b; 3 - a - b) = (a, 2; 1 - a)$  and  $\sigma = (-2, -a; 4 - a)$ . Theorem 5.2 then implies that

$$\begin{aligned} & \overline{H}^i(X_0(Q), \omega(a, 2)_{\mathbb{C}}) \\ & \subset \bigoplus_{\pi \in \mathcal{A}_{(2)}(G)} ((\pi^\infty)^{K_0(Q)} \otimes H^i(\mathrm{Lie} P^-, K^h; \pi_\infty \otimes V_\sigma))^{\oplus m_{(2)}(\pi)}, \end{aligned}$$

where  $m_{(2)}(\pi)$  denotes the multiplicity of  $\pi$  in  $\mathcal{A}_{(2)}(G)$ . Fix a degree  $i \in \{0, 1\}$ , and let  $\pi \in \mathcal{A}_{(2)}(G)$  be such that  $\pi$  contributes to  $H^i(X_0(Q), \omega(a, 2))_{\mathfrak{m}} \otimes \mathbb{C}$  under the above inclusion (for some embedding  $K \hookrightarrow \mathbb{C}$ ). Let  $\tilde{\pi}$  denote the transfer of  $\pi$  to  $\mathrm{GL}_4(\mathbb{A})$  under the Classification Theorem of [3]. Then, by Remark 5.12, the infinitesimal character of  $\tilde{\pi}_\infty$  is  $\chi_{(0, 0-(a-1), -(a-1))+3/2(1, 1, 1, 1)}$ . Let  $\chi_\pi$  denote the central character of  $\pi$ .

The representation  $\tilde{\pi}$  falls into one of six classes (a)–(f) given in [3, Section 5]. We show now that we can rule out all classes other than class (a). In cases (e) and (f),  $\tilde{\pi}$  is an isobaric sum of idèle class characters. In case (d),  $\tilde{\pi}$  is of the form  $\lambda \cdot |\cdot|^{1/2} \boxplus \lambda \cdot |\cdot|^{-1/2} \boxplus \mu$  where  $\lambda$  is an idèle class character and  $\mu$  is a cuspidal automorphic representation of  $\mathrm{GL}_2(\mathbb{A})$  such that its central character  $\chi_\mu$  satisfies  $\chi_\mu = \lambda^2 = \chi_\pi$ . Considering the infinitesimal character of  $\tilde{\pi}_\infty$ , we see that we must have  $a = 2$  and  $\mu$  must correspond to a classical modular eigenform of weight 2. In case (c), there is a cuspidal automorphic representation  $\mu$  of orthogonal type of  $\mathrm{GL}_2(\mathbb{A})$  such that  $\tilde{\pi} = \mu \cdot |\cdot|^{1/2} \boxplus \mu \cdot |\cdot|^{-1/2}$ . Being of orthogonal type means that  $\mu$  is induced from a quadratic extension of  $\mathbb{Q}$ . In case (b),  $\tilde{\pi} = \mu_1 \boxplus \mu_2$ , where the  $\mu_i$ 's are distinct cuspidal automorphic representations of  $\mathrm{GL}_2(\mathbb{A})$  with  $\chi_{\mu_1} = \chi_{\mu_2} = \chi_\pi$ . Considering the infinitesimal character of  $\tilde{\pi}_\infty$  and the fact that the  $\mu_i$ 's have the same central character, it follows that the  $\mu_i$ 's are both associated to classical modular eigenforms of weight  $a$ . Thus, in all cases (b)–(f), we can associate a compatible

family of reducible  $l$ -adic Galois representations to  $\tilde{\pi}$ . This contradicts the fact that  $\mathfrak{m}$  is non-Eisenstein.

The only remaining case is case (a) where  $\tilde{\pi}$  is a cuspidal automorphic representation of  $\mathrm{GL}_4(\mathbb{A})$  that is  $\chi_\pi$ -self dual. By Clozel's purity lemma (see [19, Lemme 4.9]),  $\tilde{\pi}_\infty$  is essentially tempered. (We thank Olivier Taïbi for pointing this out to us.) It follows that  $\pi_\infty$  is also essentially tempered, since its  $L$ -parameter is essentially bounded. Then by Theorem 5.6(3),  $\pi_\infty$  is the limit of discrete series representation  $\pi(\lambda, C_i)$ , where  $\lambda = (a-1, 0; 4-a)$ . Furthermore, by Wallach [55, Theorem 2.3], it follows that  $\pi$  is cuspidal.

By Theorem 5.2(1), the cuspidal cohomology  $\mathcal{H}_{\mathrm{cusp}, \sigma}^i$  maps injectively to the interior cohomology

$$\begin{aligned} & \overline{H}^i(X_0(Q), \omega(a, 2)_C)_{\mathrm{cusp}} \\ & \cong \bigoplus_{\pi \in \mathcal{A}_0(G)} ((\pi^\infty)^{K_0(Q)} \otimes H^i(\mathrm{Lie} P^-, K^h; \pi_\infty \otimes V_\sigma))^{\oplus m_0(\pi)}, \end{aligned}$$

where  $m_0(\pi)$  is the multiplicity of  $\pi$  in  $\mathcal{A}_0(G)$ . Thus, at this point, we can prove that the dimensions

$$\dim_K \overline{H}^j(X_0(Q), \omega(a, 2))_{\mathfrak{m}} \otimes K$$

are equal for  $j = 0, 1$  if we can establish that

- (1) the spaces  $H^j(\mathrm{Lie} P^-, K^h; \pi(\lambda, C_j) \otimes V_\sigma)$  have the same dimension for  $j = 0, 1$ ;
- (2) the representation  $\pi' = \pi^\infty \otimes \pi(\lambda, C_{1-i})$  also lies in  $\mathcal{A}_{(2)}(G)$ ;
- (3) the multiplicities  $m_0(\pi)$ ,  $m_{(2)}(\pi)$ ,  $m_0(\pi')$ , and  $m_{(2)}(\pi')$  are all equal.

The first point follows from [37, Theorem 3.4] which says that both spaces are 1-dimensional. The second point follows from [3]. Indeed, since  $\pi(\lambda, C_i)$  is essentially tempered, the local packet  $\Pi_{\psi_\infty}$  (where  $\psi = \tilde{\pi} \boxplus 1$ , in the notation of [3]) is in fact an L-packet by [55, Theorem 2.1]. Furthermore, it consists of the pair of representations  $\{\pi(\lambda, C_0), \pi(\lambda, C_1)\}$  (see [55, Section 3.1]). Since the group  $\mathfrak{S}_\psi$  is trivial in Case (a) of [3], it then follows from part (ii) of [3, Classification Theorem] that  $\pi'$  is also automorphic. Finally, for the third point, the theorem of Wallach quoted above implies that  $\pi$  and  $\pi'$  are both cuspidal. Part (iii) of [3, Classification Theorem] then implies that each of the multiplicities in point (3) is 1. We have thus shown that

$$\begin{aligned} & \dim_K H^1(X_0(Q), \omega(a, 2)(-\infty))_{\mathfrak{m}} \otimes K \\ & = \dim_K H^0(X_0(Q), \omega(a, 2)(-\infty))_{\mathfrak{m}} \otimes K, \end{aligned}$$

as required. □

### 8. $q$ -expansions of Siegel modular forms

As in Section 5.2, let  $S$  and  $Q$  be finite sets of primes of  $\mathbf{Q}$  which are disjoint and do not contain  $p$ . We allow the possibility that  $Q = \emptyset$ . We let  $K$  and  $K_i(Q)$  be open compact subgroups of  $\mathrm{GSp}_4(\mathbb{A}^\infty)$  as in Section 5.2, and we let  $X = X_K$  and  $X_i(Q) = X_{K_i(Q)}$  be the corresponding Siegel 3-folds, with open subspaces  $Y$  and  $Y_i(Q)$ , all defined over  $\mathcal{O}$ .

#### 8.1. $q$ -expansions of Siegel modular forms

(For more background and details on the results quoted in this section, see [67, Section 3.1].) Recall that  $Y_1(Q)$  has good reduction at  $p$ . Let  $R$  be an  $\mathcal{O}$ -module (we will exclusively be interested in the case when either  $R = \mathcal{O}/\varpi^n$  for some  $n$ , or when  $R = K/\mathcal{O}$ ). Let  $\sigma = (j, k) \in X^*(T)_M^+$  be a weight, and associate to  $\sigma$  the representation

$$U = (\mathrm{Sym}^{j-k}(\mathcal{O}^2) \otimes_{\mathcal{O}} \det(\mathcal{O}^2)^{\otimes k}) \otimes_{\mathcal{O}} R$$

of  $\mathrm{GL}_2$  over  $R$ . Associated to  $\sigma$ , we also have the vector bundle  $\mathcal{W}_\sigma = \omega(j, k)$ . There is a  $q$ -expansion map

$$H^0(Y_1(Q), \omega(j, k)_R) \rightarrow R[[q, q', \zeta]][[\zeta^{-1}]] \otimes_R U.$$

#### THEOREM 8.1

*The  $q$ -expansion map is injective.*

*Proof*

This is a standard fact (see, e.g., [67, Proposition 3.2]). □

#### 8.2. Explicit formulas

Let  $L$  be the product of the primes in  $S$  and  $Q$ , so that  $X_1(Q)$  has good reduction outside  $L$ . Let  $R$  be a  $\mathbf{Z}_p$ -module and thus a  $\mathbf{Z}[1/L]$ -algebra. Any  $F \in H^0(X_1(Q), \mathcal{W}_{\sigma, R})$  has a “ $q$ -expansion”

$$F = \sum_{\mathcal{X}} a(F, Q) q^{\mathcal{Q}},$$

where  $\mathcal{X}$  denotes the  $2 \times 2$  positive semidefinite matrices which take on  $\mathbf{Z}[1/L]$ -integral arguments for integral vectors, or equivalently,

$$\mathcal{X} = \begin{pmatrix} m & \frac{1}{2}r \\ \frac{1}{2}r & n \end{pmatrix}, \quad m, n, r \in \mathbf{Z}[1/L].$$

The set  $\mathcal{X}$  is naturally a subset of  $M_2(\mathbf{Q})$ . The group  $\mathrm{GL}_2(\mathbf{Q})$  acts on  $M_2(\mathbf{Q})$  by the formula



$$M.Q := (\det M)^{-1}MQM^T,$$

where the right-hand side is multiplication. We may naturally extend the definition of  $a(F, Q)$  for  $Q \in M_2(\mathbf{Q})$  by setting  $a(F, Q) = 0$  for all  $Q$  not in  $\mathcal{X}$ . In any  $q$ -expansion, the coefficients  $a(F, Q)$  will also vanish unless the denominators occurring in  $Q$  are bounded by some fixed power of  $L$  which depends only on the level structure. (Since our arguments in this section are all  $p$ -adic, there is little harm in imagining that  $L = 1$ .) Let  $V = \mathcal{O}^2$  be the standard representation of  $\mathrm{SL}_2(\mathbf{Z})$  over  $\mathcal{O}$ . The elements  $a(F, Q)$  are elements of the representation  $U$ , where, if  $\sigma$  has weight  $(j, k)$ , then

$$U = \mathrm{Sym}^{j-k}(V) \otimes R.$$

Let  $\rho : \mathrm{SL}_2(\mathbf{Z}) \rightarrow \mathrm{GL}(U)$  denote the corresponding representation. The representation  $\rho$  extends to a homomorphism from  $M_2(\mathbf{Z})$  to  $\mathrm{End}(U)$  over  $R$  which we denote by  $\rho$ , where once more  $\rho$  depends only on  $j - k$  and (more relevantly) preserves integrality. We may write the  $q$ -expansion of a form  $F$  as

$$F = \sum_{\substack{n, m \geq 0 \\ r^2 - 4mn \leq 0}} a_F(n, r, m) q^n \zeta^r q'^m,$$

where  $a_F(n, r, m) = a(F, Q)$  satisfies, for  $M \in \overline{\Gamma} \subset \mathrm{SL}_2(\mathbf{Z})$ , the equality

$$a(F, M.Q) = \rho(M)a(F, Q).$$

Here  $\overline{\Gamma}$  is the congruence subgroup of  $\mathrm{SL}_2(\mathbf{Z})$  defined in [67, p. 807]; since we are working at spherical level at  $p$ , the group  $\overline{\Gamma}$  has level prime to  $p$ . (It will do the reader little harm to pretend that  $\overline{\Gamma}$  is just  $\mathrm{SL}_2(\mathbf{Z})$ .)

*Remark 8.2*

We will assume that either  $j \geq 4$  or  $j = k = 2$ . Since we are most interested in representations with similitude character  $\nu$  equal to  $\epsilon^{j+k-3}$ , the oddness condition forces the congruence  $j \equiv k \pmod{2}$ , and so if  $j > k \geq 2$ , then  $j \geq 4$ . In cases (coming from Taylor–Wiles primes) where there is nontrivial Nebentypus character at the auxiliary primes  $q \mid Q$ , we may twist (at the cost of increasing the level at  $Q$ ) to force the Nebentypus character to be trivial. The only change this has is to make the  $q$ -expansions below less unpleasant—the addition of a Nebentypus character only introduces a notational difficulty. We note, however, that with nontrivial Nebentypus character the case of weight  $(j, k) = (3, 2)$  is possible, but our arguments would not cover this case.

### 8.3. Hecke operators at $p$

Since we will exclusively be interested in Hecke operators at  $p$ , we drop the subscript  $p$  from the notation. Similarly, we drop the subscript 1, and so  $T_{p,1}$  and  $U_{p,1}$  are denoted  $T$  and  $U$ , whereas  $T_{p,2}$  and  $U_{p,2}$  are denoted  $T_2$  and  $U_2$ , respectively. One has the following explicit description of the Hecke operator  $T$ .

#### LEMMA 8.3

*In weight  $\sigma = (j, k)$  there is an identity of formal operators  $T = U + p^{k-2}Z + p^{k+j-3}V$ , where  $U$ ,  $Z$ , and  $V$  preserve formal integral  $q$ -expansions, and such that the following identities hold:*

$$\begin{aligned} a(UF, Q) &= a(F, pQ), \\ a(ZF, Q) &= \sum_{\mathfrak{S}} \rho(M) a(F, M^{-1} \cdot Q). \end{aligned}$$

*Here  $\mathfrak{S}$  denotes (any) set of representatives in  $M_2(\mathbf{Z})$  for the left coset decomposition of*

$$\overline{\Gamma} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \overline{\Gamma}.$$

*Moreover,  $a(F, S^{-1}Q) = 0$  unless  $S^{-1}Q$  is a  $p$ -integral binary quadratic form.*

Note that the coset decomposition of  $\overline{\Gamma} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \overline{\Gamma}$  for a congruence subgroup  $\overline{\Gamma}$  prime to  $p$  is essentially the same as the coset decomposition of  $\mathrm{SL}_2(\mathbf{Z}) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \mathrm{SL}_2(\mathbf{Z})$ . These formulas are well known (see, e.g., [18, Proposition 10.2]). To compare our formula with [18], note that we have normalized the matrices in  $\mathfrak{S}$  to be integral of determinant  $p$ , and absorbed the action of the determinant into the coefficient (since we are concerned here with issues of  $p$ -integrality). We have a similar description of  $T_2$  which can be obtained by a laborious computation (following the arguments of [2, Sections 3.2 and 3.3]).

#### LEMMA 8.4

*In weight  $\sigma = (j, k)$  there is an identity of formal operators  $T_2 = p^{k+j-6}U_2 + p^{k-3}Z_2 + p^{2k+j-6}V_2$ , where  $U_2$ ,  $Z_2$ , and  $V_2$  preserve formal integral  $q$ -expansions, and the following identities hold:*

$$a(Z_2F, Q) = \sum_{\mathfrak{S}} \rho(M) a(F, M^{-1} \cdot pQ),$$

*where  $\mathfrak{S}$  is as in the description of  $Z$  in Lemma 8.3. If  $Q \not\equiv 0 \pmod{p}$ , then*

$$\begin{aligned} a(U_2F, Q) &= \left(-1 + p\left(\frac{\det(Q)}{p}\right)\right)a(Q) \\ &= \left(-1 + p\left(\frac{r^2 - 4mn}{p}\right)\right)a(Q). \end{aligned}$$

If  $Q \equiv 0 \pmod{p}$ , then

$$a(U_2F, Q) = (-1 + p^2)a(Q).$$

For those wanting a more explicit description, note that in weight  $(k, k)$  we have the possibly more familiar identities

$$\begin{aligned} a(ZF, n, r, m) &= a(pn, r, m/p) + \sum_{0 \leq \alpha < p} a((n + \alpha r + \alpha^2 m)/p, r + 2m\alpha, pm), \\ a(Z_2F, n, r, m) &= a(p^2n, pr, m) + \sum_{0 \leq \alpha < p} a((n + \alpha r + \alpha^2 m), p(r + 2m\alpha), p^2m). \end{aligned}$$

Note also that there is a formal identity  $Z_2 = UZ$ .

#### Definition 8.5

Let  $X_2$  denote the formal operator on  $q$ -expansions such that

$$U_2 = -1 + p \cdot X_2.$$

Explicitly, if  $Q \not\equiv 0 \pmod{p}$ , then  $a(X_2F, Q) = a(F, Q)$  times  $(D/p)$ , where  $D$  is the determinant of the quadratic form associated to  $Q$ , and  $(D/p)$  is the Legendre symbol. If  $Q \equiv 0 \pmod{p}$ , then  $a(X_2F, Q) = pa(F, Q)$ . In all cases, we see that  $a(X_2F, Q) = (D/p)a(F, Q) \pmod{p}$ .

#### LEMMA 8.6

Over  $k = \mathcal{O}/\varpi$ , we have  $Z_2X_2 = 0$ .

#### Proof

We have  $a(X_2F, Q) = 0$  if  $\det(Q) \equiv 0 \pmod{p}$ , but  $a(Z_2F, Q)$  is a sum over terms of the form  $a(F, R)$  with  $\det(R) = 0$ .  $\square$

#### Definition 8.7

A binary quadratic form  $Q$  is  $p$ -primitive if it is not of the form  $pR$  for a  $p$ -integral form  $R$ .

#### 8.4. Hecke operators on forms in characteristic $p$

Let  $Q_2 = (p \cdot T_2 + (p + p^3)S)p^{2-k}$ .

## LEMMA 8.8

*There is an action of  $T$  and  $Q_2$  on  $H^0(X_1(Q), \omega(j, k)_{K/\mathcal{O}})$  which commutes with the other Hecke operators and acts on  $q$ -expansions via the above formula.*

*Proof*

The argument is very similar to [35, Proposition 4.1]. It suffices to prove the result with coefficients in  $\mathcal{O}/\varpi^m$ . The natural approach to defining these operators is using correspondences, as for modular curves. There are two issues which arise. The first is that the projection maps from the Siegel modular varieties with appropriate parahoric level structures are not finite over  $X$ . The second is that the definition involving correspondences is some power of  $p$  times the actual Hecke operator of interest. A general approach to resolving these questions has been recently found by Pilloni [56], who constructs all the operators used in this paper. More importantly, his method also allows one to give an action of these operators on higher coherent cohomology as well. We use a more pedestrian approach. We can resolve the normalization issue by using the  $q$ -expansion principle. The first issue is more subtle. The geometric maps involved are certainly proper; the failure of finiteness is thus a failure of quasifiniteness. The source of quasifiniteness arises from the fact that the kernel of Frobenius of an abelian surface  $A$  could (for example) equal  $\alpha_p \times \alpha_p$ , which contains “too many” subgroup schemes of type  $\alpha_p$ . On the other hand, this issue does not arise over the ordinary locus nor over the larger almost ordinary locus consisting of abelian surfaces (those with  $p$  rank at least 1) where subgroup schemes such as  $\alpha_p \times \alpha_p$  cannot occur. This shows how to resolve the issue by the following ad hoc method. By Hartogs’s lemma, it suffices to construct  $T$  over the global sections of a subvariety  $X' \subset X$  whose complement has codimension at least 2. In particular, we may replace  $X$  by the moduli space of almost ordinary abelian surfaces for which the corresponding maps are indeed finite. Implicit in this argument is a verification that the formulas above (in Lemmas 8.3 and 8.4) preserve integrality—for  $Q_2$  this is verified in Lemma 8.12 below.  $\square$

Note that this argument is not sufficient to construct these operators on

$$H^1(X_1(Q), \omega(j, k)_{K/\mathcal{O}});$$

however, we have no need to consider the action of Hecke operators at  $p$  on these spaces.

We also need to use various properties of theta operators. We begin by recalling their basic properties.

## PROPOSITION 8.9

Let  $p > 3$ , let  $j - 2 \geq k \geq 2$ , and let  $p - 2 > j - k$ .

(1) There is a map

$$\begin{aligned} \Theta : H^0(X_1(Q), \omega(k, k)_{\mathcal{O}/\varpi^m}) \\ \rightarrow H^0(X_1(Q), \omega(k + p + 1, k + p + 1)_{\mathcal{O}/\varpi^m}) \end{aligned}$$

whose action on  $q$ -expansions is given by

$$\Theta \sum a_Q q^Q = \sum \det(Q) a_Q q^Q.$$

(2) There is a map

$$\begin{aligned} \theta_1 : H^0(X_1(Q), \omega(j, k)_{\mathcal{O}/\varpi^m}) \\ \rightarrow H^0(X_1(Q), \omega(j + p - 1, k + p + 1)_{\mathcal{O}/\varpi^m}) \end{aligned}$$

whose action on  $q$ -expansions is given by

$$\theta_1 \sum a_Q q^Q = \sum \det(Q) \text{con}(a_Q \otimes Q) q^Q,$$

where  $\text{con} : \text{Sym}^{j-k} \otimes \text{Sym}^2 \rightarrow \text{Sym}^{j-k-2}$  is the natural  $\text{SL}_2(\mathbf{Z})$ -equivariant projection.

*Proof*

The operator  $\Theta$  is defined in [75, Proposition 3.9], and the operator  $\theta_1$  is defined in [75, Proposition 3.12].  $\square$

(Some of these maps were also considered in [27]). The main results we need concerning these operators are given by the next two theorems.

## THEOREM 8.10

Let  $p > 3$ , let  $p + 1 \geq k$ , and assume that  $p \nmid k(2k - 1)$ —so in particular  $k = 2$  and  $k = p + 1$  are admissible values of  $k$ . Then the map

$$\Theta : H^0(X_1(Q), \omega(k, k)_{\mathcal{O}/\varpi^m}) \rightarrow H^0(X_1(Q), \omega(k + p + 1, k + p + 1)_{\mathcal{O}/\varpi^m})$$

is injective. In particular, if  $\Theta F = 0$ , then we must have  $F = 0$ .

*Proof*

We may immediately reduce to the case  $m = 1$  and  $\mathcal{O}/\varpi = k$ . Suppose that  $F$  lies in the kernel, so  $\Theta F = 0$ . After possibly replacing  $(k, k)$  by  $(k - (p - 1),$

$k - (p - 1)$ ), we may assume that  $F$  is not divisible by the Hasse invariant. Following [75, Theorem 4.7], it suffices to show that  $F$  is not zero on the superspecial locus if it is not divisible by the Hasse invariant. Hence  $F$  has nontrivial specialization to the  $p$ -rank 1 strata. The supersingular locus on this strata is a Cartier divisor cut out by a section of  $\omega^{(p^2-1)/2}$  for  $p > 2$ , so since  $2k < p^2 - 1$  (for  $p > 3$ ), the restriction of  $F$  is nonzero on the supersingular locus. (That the supersingular locus is a Cartier divisor inside the  $p$ -rank 1 locus when  $p > 2$  was proved by Koblitz [50, p. 193]. The exact order of vanishing can also be found in [70, Theorem 2.4].) Finally, each irreducible component of the supersingular locus is a copy of  $\mathbf{P}^1$  with  $p^2 + 1$  superspecial points on it. Moreover, the line bundle  $\omega$  restricts to  $\mathcal{O}(p - 1)$  on each of these  $\mathbf{P}^1$ 's. Hence the restriction to the superspecial points is injective as long as  $k(p - 1) \leq p^2 + 1$ , which holds for  $k \leq p + 1$ .  $\square$

We also require a related result for nonparallel weight.

THEOREM 8.11

Let  $p - 1 > j \geq 4$ . The map

$$\theta_1 : H^0(X_1(Q), \omega(j, 2)_{\mathcal{O}/\varpi^m}) \rightarrow H^0(X_1(Q), \omega(j + p - 1, p + 3)_{\mathcal{O}/\varpi^m})$$

is injective.

*Proof*

It suffices to work over  $k = \mathcal{O}/\varpi$ . Suppose that  $\theta_1 F = 0$ , and suppose that  $F$  is nonzero after restriction to the superspecial locus. Then the result follows directly from [75, Theorem 3.20]. As stated, the result does not apply in weight  $(6, 2)$ , although the same argument works in this weight, provided that one may assume (in the notation of [75, Theorem 3.20]) that  $F_2|_X \neq 0$ , which can be achieved under the action of  $\overline{\Gamma} \subset \mathrm{SL}_2(\mathbf{Z})$  for  $j < p - 1$ , since the level of  $\overline{\Gamma}$  is prime to  $p$  and so surjects onto  $\mathrm{SL}_2(\mathbf{F}_p)$ . The corresponding representation of  $\mathrm{SL}_2(\mathbf{F}_p)$  is irreducible, and thus there exists an element which applied to  $F$  has  $F_i|_X \neq 0$  for any fixed choice of  $i$ . Hence it remains to show that the restriction of  $F$  to the superspecial locus is nonzero. Let  $X = X_1(Q)$ , and denote the rank 1 strata (resp., the supersingular locus, the superspecial locus) by  $Y$ ,  $Z$ , and  $S$ , respectively. We are assuming that the restriction of  $F$  to  $Y$  is nonzero. Suppose that the restriction of  $F$  to  $Z$  is zero. There is an exact sequence

$$0 \rightarrow H^0(Y, \omega(j, 2)_k \otimes \omega^{-m}) \rightarrow H^0(Y, \omega(j, 2)_k) \rightarrow H^0(Z, \omega(j, 2)_k),$$

where  $m = (p^2 - 1)/2$ . If  $F$  restricts to zero, then we obtain a nonzero class in the first group. Yet there is also a sequence

$$\begin{aligned} H^0(X, \omega(j, 2)_k \otimes \omega^{-m}) \\ \rightarrow H^0(Y, \omega(j, 2)_k \otimes \omega^{-m}) \rightarrow H^1(X, \omega(j, 2)_k \otimes \omega^{-m-(p-1)}). \end{aligned}$$

The first term vanishes. To see that the final term vanishes, we use the fact that Serre duality shows that the last term is dual to

$$H^2(X, \omega(m+p, m+p+2-j)_k(-\infty)),$$

which vanishes by Theorem 5.1. We now have to establish nonvanishing from  $Z$  to  $S$ . The restriction of the Hodge bundle to any  $\mathbf{P}^1$  on  $Z$  is  $\mathcal{O}(-1) \oplus \mathcal{O}(p)$ . Hence we need to show that no class in

$$H^0(\mathbf{P}^1, \text{Sym}^{j-2}(\mathcal{O}(-1) \oplus \mathcal{O}(p)) \otimes \mathcal{O}(2(p-1)))$$

can vanish at  $p^2 + 1$  points. This is valid as long as

$$jp - 2 = (j-2)p + 2(p-1) \leq p^2 + 1,$$

which holds provided that  $j \leq p$ . □

### 8.5. Relationship between Hecke eigenvalues and crystalline Frobenius

Suppose that  $F$  is a cuspidal eigenform of weight  $\sigma = (j, k)$  of level prime to  $p$ , and let  $r : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\overline{\mathbf{Q}}_p)$  be the associated Galois representation. One expects (and knows in regular weights; see Theorem 6.13) that  $r$  is crystalline at  $p$  and that crystalline Frobenius has eigenvalues which are the roots of the polynomial

$$X^4 - \lambda X^3 + (p\mu + (p^3 + p)p^{k+j-6})X^2 - \lambda p^{k+j-3}X + p^{2k+2j-6},$$

where  $\lambda$  is the eigenvalue of  $T$  and  $\mu$  is the eigenvalue of  $T_2$ . We may write the eigenvalues of this polynomial as

$$\alpha, \beta p^{k-2}, \beta^{-1} p^{j-1}, \alpha^{-1} p^{k+j-3},$$

where  $\alpha$  and  $\beta$  have nonnegative  $p$ -adic valuation. That means that the coefficient of crystalline Frobenius should have characteristic polynomial

$$X^4 - (\alpha + \dots) + (\alpha\beta p^{k-2} + O(p^{k-1}))X^2 + \dots$$

On the other hand, we know that the coefficient of  $X^2$  should be

$$p^{k-2}Q_2 := p \cdot T_2 + (p + p^3)S,$$

where the operator  $Q_2$  is defined by this formula. In particular, the eigenvalues of this operator ( $Q_2$ ) should all be integral.

LEMMA 8.12

Let  $\sigma = (j, k)$  with  $j \geq k \geq 2$ . If  $(j, k) \neq (2, 2)$ , then there is a congruence of operators on formal  $q$ -expansions

$$Q_2 = (p \cdot T_2 + (p + p^3)S)p^{2-k} \equiv Z_2 \pmod{p}.$$

In particular, if  $F$  is an ordinary form of regular weight  $\sigma$  with crystalline eigenvalues as above, the eigenvalue of  $Z_2$  is  $\alpha\beta \pmod{p}$ . If  $\sigma = (2, 2)$ , then there is a congruence

$$Q_2 = (p \cdot T_2 + (p + p^3)S)p^{2-k} \equiv Z_2 + X_2 \pmod{p}.$$

*Proof*

The operator  $S$  acts by a scalar which is equal to  $p^{j+k-6}$ . Note that

$$p^3 \cdot p^{j+k-6} \cdot p^{2-k} \equiv 0 \pmod{p}.$$

Thus we can ignore the  $p^3S$  term above. We have

$$\begin{aligned} & (p \cdot T_2 + (p + p^3)S)p^{2-k} \\ &= p^{3-k}(p^{j+k-6}U_2 + p^{k-3}Z_2 + p^{j+2k-6}V_2) + p^{j-3} \pmod{p} \\ &= p^{j-3}U_2 + Z_2 + p^{j+k-3}V_2 + p^{j-3} \pmod{p} \\ &= -p^{j-3} + p^{j-2}X_2 + Z_2 + p^{j+k-3}V_2 + p^{j-3} \pmod{p} \\ &= p^{j-2}X_2 + Z_2 \pmod{p}, \end{aligned}$$

and we are done. □

### 8.6. The main theorem on $q$ -expansions

Our main theorem is as follows (we use the notation of Section 6.4).

THEOREM 8.13

Let  $\sigma = (j, 2)$  for some  $p - 1 > j \geq 2$ . Assume that  $\bar{r}$  is as in Assumption 6.16. Assume, moreover, that

$$\alpha\beta(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2\beta^2 - 1) \neq 0.$$

Let  $\mathfrak{m}$  denote the corresponding ideal of the Hecke algebra away from  $p$ . Let  $A$  denote a nontrivial power of the Hasse invariant of weight  $k$ . Then the composite map



$$\begin{array}{ccc}
 H^0(X_1(Q), \omega(j, 2)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\alpha, \beta} & \xrightarrow{A} & H^0(X_1(Q), \omega(j+k, 2+k)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}} \\
 & & \pi_{\beta} \downarrow \\
 & & H^0(X_1(Q), \omega(j+k, 2+k)_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}^{\beta}
 \end{array}$$

is injective, where  $\pi_{\beta}$  denotes the projection onto the summand where  $U - \beta$  and  $Q_2 - \alpha\beta$  (equivalently,  $Z_2 - \alpha\beta$ ) are nilpotent.

Note that, by symmetry, the same result holds with  $\beta$  replaced by  $\alpha$ . Before beginning the proof of this theorem, we first prove a much easier analogue for  $\mathrm{GL}(2)$ .

LEMMA 8.14

Let  $X_1(N)$  denote the modular curve, and let  $\bar{\rho}: G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_p)$  be a modular representation of level  $N$  and weight 1 over  $\mathbf{F}_p$  such that  $\bar{\rho}(\mathrm{Frob}_p)$  has eigenvalues  $\alpha$  and  $\beta$ . Let  $\mathfrak{m}$  denote the corresponding ideal of the Hecke algebra away from  $p$ . Assume that

$$\alpha - \beta \neq 0.$$

If  $A$  denotes a suitable power of the Hasse invariant of weight  $k$ , then the composite map

$$\begin{array}{ccc}
 H^0(X_1(N), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}} & \xrightarrow{A} & H^0(X_1(N), \omega_{\mathcal{O}/\varpi^m}^{k+1})_{\mathfrak{m}} \\
 & & \pi_{\beta} \downarrow \\
 & & H^0(X_1(N), \omega_{\mathcal{O}/\varpi^m}^{k+1})_{\mathfrak{m}}^{\beta}
 \end{array}$$

is injective, where  $\pi_{\beta}$  denotes the projection onto the quotient of homology where  $U - \beta$  is nilpotent.

In both results, all of the corresponding maps are equivariant with respect to Hecke operators away from  $p$ . It suffices to show that the image of the  $\mathbf{T}$ -socle maps injectively, and hence we may work with coefficients over a finite field  $k = \mathcal{O}/\varpi$  of characteristic  $p$ .

*Proof of Lemma 8.14*

Let  $M = H^0(X_1(N), \omega_{\mathcal{O}/\varpi^m})_{\mathfrak{m}}$ , and let  $N = H^0(X_1(N), \omega_{\mathcal{O}/\varpi^m}^{k+1})_{\mathfrak{m}}$ . The map  $M \rightarrow N$  is certainly injective, as can be seen by the  $q$ -expansion principle (the map

is the identity on  $q$ -expansions). Let  $U$  denote the action of  $T$  on  $N$ . Then  $U$  satisfies the polynomial  $U^2 - TU + \langle p \rangle = 0$  on the image of  $M$ , and so  $M$  lies inside the ordinary subspace of  $N$ , and so inside  $N_\alpha \oplus N_\beta$ , where  $N_\gamma$  is the factor of  $N$  on which  $(U - \gamma)$  is nilpotent. We have operators  $U$  and  $V$  defined by the formulas

$$U\left(\sum a_n q^n\right) = \sum a_{np} q^n, \quad V\left(\sum a_n q^n\right) = \sum a_n q^{np},$$

and  $T = U + \langle p \rangle V$  in weight 1, whereas  $T = U$  in higher weight. The projection operator

$$\pi_\beta : N_\alpha \oplus N_\beta \rightarrow N_\beta$$

is given by  $\pi_\beta = (U - \alpha)^m$  for some integer  $m$ . Suppose that  $F \in M$  satisfies  $\pi_\beta(F) = 0$ . We have the identity  $UVF = F$ , and we may reduce to the case that  $\langle p \rangle F = \alpha\beta F$ . We are assuming that  $F = F_\alpha \in N_\alpha$ . Let us write

$$(U - \alpha)F_\alpha = G_\alpha \implies UF_\alpha = \alpha F_\alpha + G_\alpha \implies \alpha U^{-1}F_\alpha = F_\alpha - U^{-1}G_\alpha.$$

Note that  $U$  is invertible on  $N_\alpha$ . Since  $TF$  also lies in  $N_\alpha \oplus N_\beta$ , we deduce that  $VF$  lies in  $N_\alpha \oplus N_\beta$ . Yet  $UVF = F \in N_\beta$ , and so  $VF \in N_\alpha$ , and moreover  $\langle p \rangle VF = \alpha\beta U^{-1}F_\alpha$ . It follows that

$$\begin{aligned} (T - \alpha - \beta)F &= (U - \alpha)F_\alpha + (\langle p \rangle V - \beta)F_\alpha \\ &= (U - \alpha)F_\alpha + \alpha\beta U^{-1}F_\alpha - \beta F_\alpha \\ &= G_\alpha + \beta F_\alpha - \beta U^{-1}G_\alpha - \beta F_\alpha \\ &= G_\alpha - \beta U^{-1}G_\alpha. \end{aligned}$$

If  $G_\alpha \neq 0$ , then the latter expression is nonzero, since applying  $U$  gives  $UG_\alpha - \beta G_\alpha$  and  $\beta \neq \alpha$ . On the other hand,  $G_\alpha$  is deeper in the filtration of  $N_\alpha$  given by

$$N_\alpha \supset (U - \alpha)N_\alpha \supset (U - \alpha)^2 N_\alpha \cdots$$

and hence, replacing  $F$  by  $(T - \alpha - \beta)F$  sufficiently many times, we may assume that  $G_\alpha = 0$ , that  $UF_\alpha = \alpha F_\alpha$ , and that  $(T - \alpha - \beta)F_\alpha = 0$ . We are thus left with a form  $F$  such that

$$TF = (\alpha + \beta)F, \quad UF = \alpha F, \quad VF = \beta F.$$

We may now achieve a contradiction based purely on a computation with formal  $q$ -expansions. For example, the identity  $VF = \beta F$  is impossible as soon as either  $\beta \neq 1$  or  $F$  is a cusp form, simply by considering the exponent of the smallest coefficient. Alternatively, a nonformal argument using properties of modular forms would be to note that  $\theta VF = 0$ , and then use the fact that  $\theta$  has no kernel in low weight (by [45]).  $\square$

A different proof of this theorem is given in [16]; the point is that the proof given here avoids any geometry. The proof below is somewhat in this spirit—using some elementary reductions, we arrive, given an element of  $\ker(\pi_\beta)$ , at a form  $F$  which is simultaneously acted upon by a collection of formal operators in a very constrained way. The identities we get are not quite enough to deduce that  $F = 0$  as formal  $q$ -expansions; however, they are enough to produce forms of low weight inside the kernel of various theta operators, which will be enough to produce a contraction by Theorems 8.11 and 8.10. No doubt (see Section 1.3) there will be better geometric replacements for this argument, so we apologize in advance for the somewhat messy approach that we present here.

As in the proof above, let us write

$$\begin{aligned} M &= H^0(X_1(Q), \omega(j, 2)_k)_m, \\ N &= H^0(X_1(Q), \omega(j + k, 2 + k)_k)_m. \end{aligned}$$

The map  $M \rightarrow N$  is certainly injective, as can be seen by the  $q$ -expansion principle (the map is the identity on  $q$ -expansions). By abuse of notation, we view  $M \subset N$  under this map. Since  $\alpha\beta \neq 0$ , the operator  $Q_2$  acts invertibly on  $M$ . Depending on the weight  $\sigma$ , the operator  $Q_2$  acts on  $M$  either as  $Z_2$  or as  $Z_2 + X_2$ .

#### LEMMA 8.15

*Assume that  $\alpha$  and  $\beta$  are as in Theorem 8.13. Suppose that  $\sigma = (j, 2)$  with  $j > 2$ . Then  $M = Q_2M = Z_2M$ , and  $M$  is a subspace of the submodule of  $N$  on which  $U$  is invertible. If  $\sigma = (2, 2)$ , then  $Z_2$  acts on  $N$ , the map  $M \rightarrow Z_2M$  is injective, and  $Z_2M \subset N$  is a subspace of the submodule of  $N$  on which  $U$  is invertible.*

#### *Proof*

In the first case, by assumption we know that  $Q_2 - \alpha\beta$  is nilpotent, and so  $Q_2$  induces an isomorphism of  $M$ . On the other hand, the operator  $Q_2$  acts via the formal operator  $Z_2$ . In weight  $\tau = (j + k, 2 + k)$ , the corresponding operator  $Q_2$  also acts via  $Z_2$ , and so we deduce that  $Q_2 - \alpha\beta$  acts on  $M \subset N$  and acts nilpotently. Yet  $Q_2$  only acts invertibly on the ordinary part of  $N$ , as can be seen by lifting to characteristic 0. Now let us consider the case of weight  $\sigma = (2, 2)$ . We have

$$M = Q_2M = (Z_2 + X_2)M.$$

Now  $Q_2$  acts in weight  $N$  by  $Z_2$ , so certainly  $Z_2M \subset N$ . Since  $Q_2$  acts by  $Z_2 + X_2$  on  $M$ , there is a commutative diagram

$$\begin{array}{ccc}
M & \longrightarrow & Z_2M \\
\downarrow & & \downarrow \\
Q_2M & \longrightarrow & Z_2(Z_2 + X_2)M = Z_2^2M
\end{array}$$

where (by Lemma 8.6) we use the fact that  $Z_2X_2 = 0$ . Since the left-hand side is an isomorphism, it follows that  $Z_2^2M = Z_2M$ , and hence that  $Z_2$  acts invertibly on  $Z_2M$ , and as in the previous argument it follows that  $Z_2$  and hence  $U$  is invertible on this space.

Hence it suffices to show that  $Z_2F \neq 0$  for any  $F \in M$ . Suppose that  $Z_2F = 0$ . Then  $Q_2F = Z_2F + X_2F = X_2F$ . Since  $Q_2F \in M$ , we have  $X_2F \in M$ . Yet then (again by Lemma 8.6) we have  $Q_2^2F = (Z_2 + X_2)X_2F = X_2^2F$ , and then  $Q_2^3F = X_2^3F = X_2F$ , and so  $Q_2F = X_2F \neq 0$  is an eigenvector of  $Q_2$  with eigenvalue  $\lambda$  satisfying  $\lambda^2 = 1$ . Yet the only generalized eigenvalue of  $Q_2$  is  $\alpha\beta$ , and by assumption  $(\alpha\beta)^2 \neq 1$ .  $\square$

(Note that this is the point in this paper which uses the assumption  $(\alpha\beta)^2 \neq 1$  rather than the weaker claim  $\alpha\beta \neq 1$  which is sufficient for arguments on the Galois side.)

LEMMA 8.16

*The operator  $U(U - \alpha)(U - \beta)$  acts nilpotently on  $N$ .*

*Proof*

This follows by lifting to characteristic 0 and noting that the only possible unit crystalline eigenvalues of Frobenius of a lift of  $\bar{r}$  are  $\alpha$  or  $\beta$  modulo  $\mathfrak{m}$ .  $\square$

LEMMA 8.17

*Suppose that the composite  $\pi_\beta : Z_2M \rightarrow N_\beta$  is not injective.*

- (1) *If  $(\sigma) = (j, 2)$  with  $j > 2$ , then there exists a nonzero form  $F = F_\alpha \in M \cap N_\alpha$  such that*

$$UF = \alpha F, \quad TF = (\alpha + \beta)F, \quad ZF = \beta F.$$

- (2) *If  $(\sigma) = (2, 2)$ , then there exists a nonzero form  $F = F_\alpha + F_0$  with  $F_\alpha \in N_\alpha$  and  $F_0 \in N_0$  such that*

$$UF_\alpha = \alpha F_\alpha, \quad TF = (\alpha + \beta)F, \quad X_2F = \alpha\beta F_0.$$

*Proof*

First note that  $TF = (U + Z)F \in M$ , and that  $UF \in N$ , so  $ZF \in N$ . Assume that  $\sigma = (j, 2)$  with  $j > 2$ . Note that  $Z_2$  commutes with  $U$ . Hence, after replacing  $F \in \ker(\pi_\beta)$  by  $(Z_2 - \alpha\beta)^m F = (Q_2 - \alpha\beta)^m F$  for sufficiently large  $m$ , we may assume that  $Z_2 F = \alpha\beta F$ . The assumption  $\pi_\beta(F) = 0$  implies that  $F = F_\alpha \in N_\alpha$ . Clearly  $UF \in N_\alpha$  also, and so  $ZF = TF - UF \in N_\alpha \oplus N_\beta$ . Yet  $Z_2 = UZ$ , so we have

$$UZF = \alpha\beta F_\alpha \Rightarrow ZF = \alpha\beta U^{-1} F_\alpha \in N_\alpha.$$

(There can be no component in  $N_\beta$  because  $U$  is invertible on that space.) Write  $(U - \alpha)F_\alpha = G_\alpha$ , so  $UF_\alpha - G_\alpha = \alpha F_\alpha$ , or

$$\alpha U^{-1} F_\alpha = F_\alpha - U^{-1} G_\alpha.$$

We infer that

$$\begin{aligned} (T - \alpha - \beta)F &= (U + Z - \alpha - \beta)F_\alpha \\ &= (U - \alpha)F_\alpha + (Z - \beta)F_\alpha \\ &= G_\alpha + \beta F_\alpha - \beta U^{-1} G_\alpha - \beta F_\alpha \\ &= G_\alpha - \beta U^{-1} G_\alpha. \end{aligned}$$

We claim that if  $G_\alpha \neq 0$ , then the last expression is nonzero. This is because  $U$  acts invertibly on  $N_\alpha$ , and applying  $U$  we get

$$U(G_\alpha - \beta U^{-1} G_\alpha) = (U - \alpha)G_\alpha + (\alpha - \beta)G_\alpha,$$

where  $(U - \alpha)G_\alpha$  has a smaller nilpotence level than  $G_\alpha$ , and  $(\alpha - \beta) \neq 0$ . In particular, replacing  $F$  by  $(T - \alpha - \beta)F$ , we may find more elements in  $M$  which also lie in the kernel of  $\pi_\beta$ , and reduce to the case where  $UF_\alpha = \alpha F_\alpha$  and  $Z_2 F_\alpha = UZ_2 F_\alpha = \alpha\beta F_\alpha$ . However, in this case, we also see that  $ZF_\alpha = \beta F_\alpha$ , and the required equalities follow.

Now suppose that  $\sigma = (2, 2)$ . Let us write  $\pi_\beta : Z_2 M \subset N_\alpha \oplus N_\beta \rightarrow N_\beta$  as  $(U - \alpha)^m$ , and so  $(U - \alpha)^m Z_2 F = 0$  for some  $F \neq 0$ . Since  $Z_2$  formally commutes with  $U$ , we also get

$$(U - \alpha)^m (Z_2^2) F = Z_2 (U - \alpha)^m Z_2 F = 0,$$

so  $Z_2$  preserves the property of  $Z_2 F$  lying in the kernel of  $\pi_\beta$ . But

$$Z_2(Z_2 + X_2)F = Z_2^2 F,$$

because  $Z_2 X_2 = 0$ . Hence, if  $Z_2 F$  lies in the kernel of  $\pi_\beta$ , then so does

$$Z_2 Q_2 F = Z_2 (Z_2 + X_2) F.$$

Hence we may repeatedly replace  $F$  by  $(Q_2 - \alpha\beta)F = (Z_2 + X_2 - \alpha\beta)F$ , and thus replace  $F$  by a form such that  $Q_2 F = \alpha\beta F$  and  $Z_2 F \in N_\alpha$ . Now, as above, we may write

$$F = F_\alpha + F_0 = (F_\alpha, 0, F_0) \in N_\alpha \oplus N_\beta \oplus N_0.$$

We are assuming that  $Q_2 F = \alpha\beta F$ , and so

$$Q_2 F = (\alpha\beta F_\alpha, 0, \alpha\beta F_0).$$

Thus we deduce that  $X_2 F = (0, 0, \alpha\beta F_0)$  and  $Z_2 F = (\alpha\beta F_\alpha, 0, 0)$ . We once more would like to use that  $T = U + Z$  implies that  $ZF \in N$ . However, we no longer know (or expect) that  $ZF$  is ordinary. However, since  $Z_2 = UZ$  and  $ZF \in N$ , we certainly deduce that

$$ZF = (U^{-1}\alpha\beta F_\alpha, 0, G_0),$$

for some  $G_0$  in the kernel of  $U$ . Our arguments are similar to those used above. We write  $(U - \alpha)F_\alpha = G_\alpha$ , so  $UF_\alpha - G_\alpha = \alpha F_\alpha$ , or

$$\alpha U^{-1} F_\alpha = F_\alpha - U^{-1} G_\alpha.$$

This implies that

$$\begin{aligned} G &:= (T - \alpha - \beta)F \\ &= (U + Z - \alpha - \beta)(F_\alpha, 0, F_0) \\ &= (\alpha F_\alpha + G_\alpha, 0, 0) + (\beta F_\alpha - \beta U^{-1} G_\alpha, 0, G_0) - (\alpha + \beta)F \\ &= (G_\alpha - \beta U^{-1} G_\alpha, 0, G_0 - (\alpha + \beta)F_0). \end{aligned}$$

The first term lies in a space where  $(U - \alpha)$  is nilpotent, but it has a smaller nilpotence level than  $F_\alpha$  by construction. Moreover, if it is equal to zero, then

$$0 = U(G_\alpha - \beta U^{-1} G_\alpha) = \alpha G_\alpha + H_\alpha - \beta G_\alpha,$$

where  $(U - \alpha)G_\alpha = H_\alpha$  has yet a higher level of nilpotence. In particular, this can equal zero only if either  $\alpha = \beta$  or  $G_\alpha = 0$ . Since we are explicitly forbidding the former, we may assume, by induction, that  $F_\alpha \neq 0$  is a  $U$ -eigenvector, and

so

$$(T - \alpha - \beta)F = (0, 0, G_0 - (\alpha + \beta)F_0).$$

This implies that  $Z_2(T - \alpha - \beta)F = 0$ , and thus (from the injectivity of  $Z_2$  in Lemma 8.15) that  $(T - \alpha - \beta)F = 0$ , or that  $F$  is a  $T$ -eigenform. The required identities follow immediately upon writing  $F = F_\alpha + F_0$ , where  $F$  is a  $T$ -eigenform,  $UF_\alpha = \alpha F_\alpha$ , and  $X_2F = \alpha\beta F_0$ .  $\square$

At this point, to prove Theorem 8.13, it suffices to show that there are no Siegel modular forms which satisfy the above identities. For example, in weights  $\sigma = (j, 2)$  with  $j > 2$ , we would like to show that there is no form  $F$  which is an eigenform for both  $T$  and  $U$ . We now examine what constraints these identities place on the Fourier coefficients of  $F$ .

*Remark 8.18 (Tripling)*

A theme of [16], following previous work of Wiese [72], was to prove that certain Galois representations were ordinary in two different ways by *doubling*, that is, mapping the form of low weight to forms of high weight in two different ways. This is also our argument in weights  $(j, 2)$  for  $j \geq 4$ . However, in weight  $(2, 2)$ , we see some new phenomena. When we pass to weight  $(p + 1, p + 1)$ , we find that the image of the space of low-weight forms generates (under the action of Hecke operators) a space whose dimension has not only doubled, but has rather tripled. Moreover, this space of oldforms contains nonzero elements in the kernel of  $Z_2$ . What this must mean is that, in weight  $(p + 1, p + 1)$ , any ordinary Galois representation coming from weight  $(2, 2)$  should have a nonordinary lift in weight  $(p + 1, p + 1)$ . This phenomena does not happen for  $GL(2)$ , since forms of weight  $p$  which are ordinary modulo  $p$  are ordinary in characteristic 0 by (boundary cases of) Fontaine–Laffaille theory. For  $GSp(4)$ , however, the Hodge–Tate weights in weight  $(p + 1, p + 1)$  are  $[0, p - 1, p, 2p - 1]$ , which are well beyond the Fontaine–Laffaille range. One might also ask what the exact relationship is between the tripling argument here in weight  $(2, 2)$  and the doubling version of [13] at Klingen level. For our purposes, this would require proving that there exists a (Hecke equivariant away from  $p$ ) injection from our space of forms  $M$  at spherical level to a space of ordinary forms (with respect to the operator denoted  $U_{\text{Kli},2}$  in [13]) at Klingen level also in weight  $(2, 2)$ . While this should certainly be true, we have not attempted to prove it.

### 8.7. Binary quadratic forms

#### Definition 8.19

We define a set with multiplicities  $\mathcal{F}(Q)$  of equivalence classes of  $p$ -integral binary quadratic forms as follows. For each  $M \in \mathcal{S}$  (with  $\mathcal{S}$  as defined in Lemma 8.3), we add  $[P]$  to  $\mathcal{F}(Q)$  if and only if there exists a  $P \in [P]$  such that  $Q = M.P$ . In particular,  $M$  contributes a class  $[P]$  if and only if  $[M^{-1}.Q]$  is  $p$ -integral.

An easy lemma shows that  $\mathcal{F}(Q)$  depends only on  $[Q]$ . A binary quadratic form defines a section of  $\mathcal{O}(2)$  on  $\mathbf{P}^1(\mathbf{F}_p)$ , the latter of which is in natural bijection to  $\mathcal{S}$  (recall that  $\mathcal{S}$  is the coset space of  $\text{diag}(1, p)$  in  $\overline{\Gamma} \subset \text{SL}_2(\mathbf{Z})$ ). We see that  $M^{-1}.Q$  is  $p$ -integral if and only if the corresponding quadratic form has a zero at the corresponding point in  $\mathbf{P}^1(\mathbf{F}_p)$ . In particular,  $\mathcal{F}(Q)$  is empty if  $Q$  does not represent zero. Moreover, the cardinality of  $\mathcal{F}(Q)$  is given by the number of zeros of  $Q$ , and is thus equal to 0, 1, or 2 if  $Q$  is  $p$ -primitive. (If  $Q$  is not  $p$ -primitive, then  $Q \equiv 0 \pmod p$  and  $\mathcal{F}(Q)$  has cardinality  $p + 1$ .)

The definition of  $\mathcal{F}(Q)$  is motivated by the following observation. There is an identity

$$a(ZF, Q) = \sum_{[P] \in \mathcal{F}(Q)} \rho(M_P) a(F, P),$$

where  $P \in [P]$  is some (any) element in  $[P]$  such that  $M_P.P = Q$  for  $M_P \in \mathcal{S}$ .

#### LEMMA 8.20

If  $[P] \in \mathcal{F}([Q])$ , then  $[Q] \in \mathcal{F}([P])$ .

#### Proof

Replacing  $Q$  by  $g.Q$  for some  $g \in \overline{\Gamma} \subset \text{SL}_2(\mathbf{Z})$ , we may assume that  $Q = M.P$  where

$$M = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}.$$

Yet then  $pM^{-1}.Q = M^{-1}.Q = P$ , and  $pM^{-1} \in \mathcal{S}$ . □

Let  $d(Q)$  denote the discriminant of  $Q$ .

#### LEMMA 8.21

Suppose that  $Q$  is  $p$ -primitive. Let  $D = d(Q)$ . Then either

- (1)  $(D/p) = -1$ , and  $\mathcal{F}([Q])$  is empty,
- (2)  $(D/p) = 0$ , and  $\mathcal{F}([Q])$  has exactly one element, or
- (3)  $(D/p) = +1$ , and  $\mathcal{F}([Q])$  has exactly two elements.



*Proof*

This follows from the fact that a  $p$ -primitive form  $Q$  has exactly 0, 1, or 2 solutions in  $\mathbf{P}(\mathbf{F}_p)$ , depending on whether  $(D/p)$  is  $-1$ ,  $0$ , or  $1$ , respectively. Note that (in the final case)  $\mathcal{F}([Q])$  may consist of the same class with multiplicity 2. This happens, for example, if  $(D/p) = 1$  and the class number of  $D$  is 1.  $\square$

In light of Lemma 8.17, to prove Theorem 8.13, it suffices to prove the following.

THEOREM 8.22

Suppose that  $F = \sum a(F, Q)q^Q$  is a Siegel modular  $q$ -expansion of weight  $\sigma = (j, 2)$  in characteristic  $p$ , where  $p - 1 > j$ .

- (1) Let  $\sigma = (j, 2)$  with  $j \geq 4$ , and suppose that  $UF = \alpha F$  and  $ZF = \beta F$  for some  $\alpha, \beta$  with  $\alpha\beta(\beta^2 - 1) \neq 0$ . Then  $F = 0$ .
- (2) Let  $\sigma = (2, 2)$ , and suppose that  $F = F_\alpha + F_0$ , where  $UF_\alpha = \alpha F_\alpha$ ,  $X_2 F = \alpha\beta F_0$ , and  $ZF = \beta F + \alpha F_0$  for some  $\alpha, \beta$  with  $\alpha\beta(\beta^2 - 1)(\alpha^2\beta^2 - 1) \neq 0$ . Then  $F = 0$ .

*Proof*

We first prove that there exists a  $Q$  with  $\det(Q) \not\equiv 0 \pmod{p}$ . In particular, in weight  $(2, 2)$ , we may also assume that  $F_0 = (\alpha\beta)^{-1}X_2 F = 0$ , and thus have the equalities

$$UF = \alpha F, \quad ZF = \beta F.$$

In fact, we may assume that these equalities hold in both cases, since we are assuming that such an equality holds in the case of nonparallel weight. If  $a(F, pP) \neq 0$ , then, since  $a(F, pP) = a(UF, P) = \alpha \cdot a(F, P)$ , we have  $a(F, P) \neq 0$ . Hence, if  $F \neq 0$ , then there exists a  $p$ -primitive form  $Q$  with  $a(F, Q) \neq 0$ . Without loss of generality, assume that  $Q$  is a  $p$ -primitive form of minimal discriminant with  $a(F, Q) \neq 0$ . By Lemma 8.21,  $\mathcal{F}(Q)$  consists of a single class  $[P]$ . It follows that

$$a(F, P) = \rho(M_Q)a(ZF, Q) = \beta \cdot \rho(M_Q)a(F, Q).$$

If  $P$  is not  $p$ -primitive, then  $P = pR$  for some  $R$ , and then  $a(F, R) \neq 0$ , contradicting the minimality of  $Q$  (note that  $P$  and  $Q$  have the same discriminant). Hence  $P$  is also  $p$ -primitive. Yet then  $\mathcal{F}(P)$  consists of a single element, which must be  $[Q]$  by Lemma 8.20. Yet then it follows that

$$\begin{aligned} \beta^2 a(F, Q) &= a(Z^2 F, Q) \\ &= \rho(M_P)a(ZF, P) \end{aligned}$$

$$\begin{aligned}
&= \rho(M_Q)\rho(M_P)a(F, Q) \\
&= \begin{cases} 0 & j > 2, \\ a(F, Q) & j = 2. \end{cases}
\end{aligned}$$

Here we use that  $P = M_Q \cdot Q = M_Q \cdot M_P \cdot P$ , and thus  $\rho(M_Q \cdot M_P) = \rho(p \cdot I)$  is the identity in weight  $(2, 2)$  and zero in higher weight. If  $j > 2$ , we are done, and if  $\sigma = (2, 2)$ , we are done since  $\beta^2 - 1 \neq 0$ .

*Remark 8.23*

As an alternative to this argument, one could use an analogue of Theorem 8.10 to show that the kernel of  $\Theta$  is trivial in low weight (but this would require formulating and then proving such a theorem for nonparallel weight).

We may therefore assume that  $a(F, Q) \neq 0$  for some  $Q$  of discriminant  $D$  prime to  $p$ .

*8.8. The case  $\sigma = (2, 2)$*

Let us now assume that  $\sigma = (2, 2)$ . The coefficient  $a(X_2F, Q)$  is equal to  $(D/p)a(F, R)$ , where  $D = D_Q$  is the discriminant of  $Q$ . Hence, since  $ZF = \beta F + \beta^{-1}X_2F$ , we deduce that, if  $(D/p) = -1$ , then

$$0 = a(ZF, Q) = \beta a(F, Q) - \beta^{-1}a(F, Q) = (\beta - \beta^{-1})a(F, Q).$$

Assuming that  $\beta^2 \neq 1$ , we deduce that  $a(F, Q) = 0$ . It follows that the only  $Q$  with  $a(F, Q) \neq 0$  has  $D = \det(Q)$  satisfying  $(D/p) = 0, 1$ . In particular, the form

$$F - X_2F \in M$$

lies in the kernel of  $\Theta$ . Yet this implies that  $F - X_2F$  is trivial by Theorem 8.10. But this implies that  $Z_2F = Z_2X_2F = 0$ , contradicting the injectivity of  $Z_2 : M \rightarrow N$  in Lemma 8.15.

*8.9. The case  $\sigma = (j, 2)$  with  $j \geq 4$*

We may assume that  $a(F, Q) \neq 0$ , where  $Q$  is  $p$ -primitive and  $D = d(Q)$  is nonzero. If  $(D/p) = -1$ , then  $a(ZF, Q) = 0$ , contradicting the nonvanishing of  $a(F, Q)$  and the identity  $ZF = \beta Z$ . Hence we may assume that  $(D/p) = 1$ . The action of  $\overline{\Gamma} \subset \mathrm{SL}_2(\mathbf{Z})$  on binary quadratic forms of discriminant  $D$  has a finite orbit which may be identified with a ray class group. The assumption on  $D$  implies that  $Q$  has exactly two zeros in  $\mathbf{P}^1(\mathbf{F}_p)$ . For either of the zeros (say  $\xi$ ), we may consider the corresponding

quadratic form

$$P = M.Q := MQM^T \cdot \det(M)^{-1},$$

where  $M$  is a representative of an element in  $\mathcal{S}$  corresponding to  $\xi$ . The class of  $P$  in the class group does not depend on the choice of representative of  $M$ . The quadratic form  $P$  also has two roots. We claim that, for one of those roots, there is a choice of representative  $N$  for the element in  $\mathcal{S}$  such that

$$N.P := NPN^T \cdot \det(N)^{-1} = Q, \quad MN = \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix}.$$

Indeed, if  $N = pM^{-1}$ , then the corresponding identity is trivially satisfied. We may view the process of applying  $Z$  dynamically as follows. The coefficient corresponding to a quadratic form  $Q$  of discriminant  $D$  with  $(D/p) = +1$  of  $ZF$  is given by a sum  $\rho(M_P)a(F, P) + \rho(M_R)a(F, R)$  for a pair of quadratic forms  $P$  and  $R$  also of the same discriminant. The ray class group corresponding to  $Q$  is partitioned by this process  $Q \rightarrow \{P, R\}$  into a finite number of cyclic orbits, on which this operation takes a binary quadratic form to its two nearest neighbors (if the orbit has fewer than two elements, this pair of neighbors may have multiplicity). Let us now consider the coefficient  $a(Z^2F, Q)$ . This consists of two pairs of two terms coming from the neighboring quadratic forms  $P$  and  $R$ , respectively. From the above, for each neighbor  $P$ , there will be a term of the form

$$\rho(M)\rho(N)a(F, Q) = \rho(MN)a(F, Q) = 0,$$

where the identity  $\rho(MN) = 0$  requires the assumption that  $j > k$ . Hence  $a(Z^2F, Q)$  will also be a sum of two terms coming from the quadratic forms of distance 2 away from  $Q$  inside its cyclic orbit. Let us consider one orbit of size  $s$ . Then, we also see, modifying  $M_s$  by an element of  $\overline{\Gamma}$  if necessary, that

$$Q = M_s M_{s-1} \dots M_1 . Q = AQA^t \cdot \det(A)^{-1},$$

where  $A = M_s M_{s-1} \dots M_1 \in M_2(\mathbf{Z})$  has  $\det(A) = p^s$ . Cycling the other way, we deduce the following.

LEMMA 8.24

*Suppose that  $F$  is a formal Siegel modular form of weight  $(j, 2)$  which is an eigenform of  $Z$  with eigenvalue  $\beta$ . Suppose that  $Q$  has discriminant  $D$  with  $(D/p) = 1$ . Then there exists an integer  $s > 0$  such that*

$$\begin{aligned} \beta^s a(F, Q) &= \rho(A)a(F, Q) + \rho(B)a(F, Q) \\ &= \text{Sym}^{j-2}[A]a(F, Q) + \text{Sym}^{j-2}[B]a(F, Q), \end{aligned}$$

where

$$AQA^t = p^s Q, \quad BQB^t = p^s Q.$$

To briefly recap, at the beginning of the proof of Theorem 8.22, we proved that we could assume that  $F$  had a nonzero coefficient  $a(F, Q)$  where  $Q$  has nonzero discriminant modulo  $p$ . If  $(D/p) = -1$ , then  $a(ZF, Q) = 0$ , which (with  $ZF = \beta F$ ) would imply that  $F = 0$ . Hence we may assume that there is a nonzero coefficient with  $(D/p) = +1$  (which we exploit below) and use the following proposition to reach the final contradiction.

PROPOSITION 8.25

*Suppose that  $F$  is a formal Siegel modular form of weight  $(j, 2)$  modulo  $p$  which is an eigenform of  $Z$  with eigenvalue  $\beta$  such that  $\beta \neq 0$ , and suppose that  $p > j - 2$ . Suppose that  $F$  has a nonzero coefficient  $a(F, Q)$  where  $(D/p) = 1$ . Then  $\theta_1 F = 0$ .*

*Proof*

The map  $\theta_1$  is induced from the contraction map

$$\text{con} : \text{Sym}^{j-2} \otimes \text{Sym}^2 \rightarrow \text{Sym}^{j-4} \otimes \det$$

(this is well defined integrally as long as  $p > j - 2$ ). In particular, we have the identity

$$\begin{aligned} a(\theta_1 Z^s F, Q) &= \text{con}(\text{Sym}^{j-2}[A]a(F, Q) \otimes Q^\vee) \\ &\quad + \text{con}(\text{Sym}^{j-2}[B]a(F, Q) \otimes Q^\vee), \end{aligned}$$

where  $\text{con}$  denotes the contraction map. We claim that  $\text{con}(\text{Sym}^{j-2}[A]x \otimes Q^\vee) = 0$  for any  $x \in \text{Sym}^{j-2}V$ , where  $V = k^2$ . Once we have this, we deduce that  $\beta^s a(\theta_1 F, Q) = a(\theta_1 Z^s F, Q) = 0$ , and since  $\beta \neq 0$ , we have  $a(\theta_1 F, Q) = 0$  and  $\theta_1 F = 0$ .

While there is probably an easy coordinate-free way to prove the required claim, it is also simple enough to do the computation explicitly by writing everything out in terms of bases. Let us write down a standard basis  $\{f_1, f_2\}$  for  $V$  and a standard basis  $\{e_1, e_2\}$  for  $V^\vee$ . To be explicit, we choose bases such that a form

$$Q = \begin{pmatrix} m & \frac{1}{2}r \\ \frac{1}{2}r & n \end{pmatrix}$$

gives rise to the element  $mf_1^2 + rf_1f_2 + nf_2^2$ , and  $Q^\vee$  gives rise to  $me_1^2 + re_1e_2 + ne_2^2$ . With respect to this choice, the contraction map on  $\text{Sym}^2 \otimes \text{Sym}^2$  (up to scalar)

corresponds to sending  $e_1^2 f_2^2$  and  $e_2^2 f_1^2$  to  $-2$  and  $e_1 f_1 e_2 f_2$  to  $1$ , and sending all other monomials to zero. As a consistency check, note that

$$\text{con}(Q \otimes Q^\vee) = r^2 - 4mn = -4 \det(Q).$$

Similarly, the contraction mapping on  $\text{Sym}^{j-2} \otimes \text{Sym}^2$  for  $p > j - 2$  satisfies

$$\text{con}(f_1^i f_2^{j-2-i} e_1^k e_2^{2-k}) = 0 \quad \text{unless } 2 \leq i + k \leq j - 4.$$

The formula  $Q = AQA^t \det(A)^{-1}$  continues to hold if we replace  $Q$  by  $M \cdot Q = MQM^t$  and  $A$  by  $MAM^{-1}$  for some invertible  $M$ . In particular, we may replace  $A$  by any integral conjugate. We consider two cases.

- (1)  $A$  has a nonzero eigenvalue mod  $p$ . In this case (by Hensel's lemma), the matrix  $A$  has an eigenvalue over  $\mathbf{Z}_p$ , and a second eigenvalue which has valuation  $s$ . In particular, after a change of basis, we may write

$$A = \begin{pmatrix} u & 0 \\ 0 & 0 \end{pmatrix} \pmod{p^s},$$

$$Q = \begin{pmatrix} m & \frac{1}{2}r \\ \frac{1}{2}r & n \end{pmatrix}.$$

The conditions  $AQA^t = \det(A)Q$  and  $\det(A) = p^s$  imply that  $n \equiv 0 \pmod{p^s}$  (multiply out and consider the bottom right entry), and thus that  $Q^\vee = me_1^2 + re_1e_2 \pmod{p}$ . But now the image of  $A$  on  $k$  is generated by  $f_1$ , and so the image of  $\text{Sym}^{j-2}[A]x$  is given by  $f_1^{j-2}$ . But this forces the contraction after tensoring with  $Q^\vee$  to be zero over  $k$ , because the only monomial which  $f_1^{j-2}$  contracts nontrivially with is  $e_2^2$ .

- (2)  $A$  is nilpotent modulo  $p$ . If  $A$  is trivial modulo  $p$ , then there is nothing to prove. On the other hand, if

$$A \equiv \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \pmod{p},$$

then once again the image of  $A$  is generated by  $f$ , and the conditions  $AQA^t = \det(A)Q$  and  $\det(A) = p^s$  imply once more that  $n \equiv 0 \pmod{p}$  (multiply out as above but now consider the top left entry), and the proof proceeds as in the previous case.

This completes the proof of the proposition.  $\square$

Combining Proposition 8.25 with Lemma 8.17 and Theorem 8.11, we obtain a contradiction, and this completes the proof of Theorem 8.13.  $\square$

### 9. Modularity lifting

The following theorem is the main result of this paper.

#### THEOREM 9.1

Let  $\bar{r} : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(k)$  be a continuous, odd, absolutely irreducible Galois representation. Suppose that  $v(\bar{r}) = \epsilon^{-(a-1)}$ , where  $p-1 > a \geq 2$ . Suppose that the following hold:

- (1) There exist units  $\alpha$  and  $\beta$  in  $k$  such that

$$\bar{r} | G_p \sim \begin{pmatrix} \lambda(\alpha) & 0 & * & * \\ 0 & \lambda(\beta) & * & * \\ 0 & 0 & v(\bar{r}) \cdot \lambda(\beta^{-1}) & 0 \\ 0 & 0 & 0 & v(\bar{r}) \cdot \lambda(\alpha^{-1}) \end{pmatrix},$$

and moreover  $(\alpha^2 - 1)(\beta^2 - 1)(\alpha^2\beta^2 - 1)(\alpha - \beta) \neq 0$ .

- (2) Let  $S(\bar{r})$  denote the set of primes of  $\mathbf{Q}$  away from  $p$  at which  $\bar{r}$  is ramified. Then for each  $x \in S(\bar{r})$ , the restriction  $\bar{r} | G_x$  falls into one of the cases of Assumption 4.3.
- (3) (Big image) The restriction  $\bar{r} | G_{\mathbf{Q}(\zeta_p)}$  has big image in the sense of Assumption 4.1.
- (4) The representation  $\bar{r}$  is Katz modular of weight  $\sigma := (a, 2) \in X^*(T)_M^+$  in the sense of Definition 6.15.
- (5) (Neatness) The representation  $\bar{r}$  satisfies Assumption 4.2.

We now introduce some notation. Let  $K \subset \mathrm{GSp}_4(\mathbb{A}^\infty)$  be the compact open subgroup defined as in the beginning of Section 6.3. Let  $X = X_K$ , and for any set of primes  $Q$  disjoint from  $S(\bar{r}) \cup \{p\}$ , let  $X_i(Q) = X_{K_i(Q)}$ . Let the Hecke algebras  $\mathbf{T}_\sigma$  and  $\mathbf{T}_\sigma^{\mathrm{an}}(Q)$  be as in Definition 5.13. The assumption that  $\bar{r}$  is Katz modular implies that there is a maximal ideal  $\mathfrak{m}_\emptyset$  of  $\mathbf{T}_\sigma$  associated to  $\bar{r}$ . The pullback of  $\mathfrak{m}_\emptyset$  to  $\mathbf{T}_\sigma^{\mathrm{an}}(Q)$  is also denoted  $\mathfrak{m}_\emptyset$ . We further assume the following.

- (6) If  $Q$  satisfies Assumption 6.12(2), then

$$H^2(X_i(Q), \omega(a, 2)(-\infty)_k)_{\mathfrak{m}_\emptyset} = \{0\}.$$

Let  $R^{\mathrm{min}}$  be the universal deformation ring classifying minimal deformations of  $\bar{r}$  in the sense of Definition 4.6 (with  $Q$  taken to be empty). Then the map

$$R^{\mathrm{min}} \rightarrow \mathbf{T}_{\sigma, \mathfrak{m}_\emptyset}^{\alpha, \beta},$$

which classifies the minimal deformation of Theorem 6.17 (with  $Q$  taken to be empty), is an isomorphism. Furthermore, the space

$$H^0(X, \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha, \beta, \vee}_{\mathfrak{m}_\emptyset}$$

is a free  $\mathbf{T}_{\sigma, \mathfrak{m}_\emptyset}^{\alpha, \beta}$ -module.

Note that, for  $p \geq a \geq 4$ , hypothesis (6) holds by Theorem 5.1.

*Proof*

To prove the theorem, we apply Proposition 3.3 as follows:

- (1) Take  $R = R^{\min}$  and  $H = H^0(X, \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha, \beta, \vee}_{\mathfrak{m}_\emptyset}$ .
- (2) Let  $q$  and the sets  $Q_N$  be as in Proposition 4.10.
- (3) The ring  $R_\infty$  is the power series ring  $\mathcal{O}[[x_1, \dots, x_{q-1}]]$ .
- (4) For each  $N \geq 1$ , we define a surjection  $R_\infty \rightarrow R$  as follows. Let  $R_{Q_N}$  denote the universal deformation ring classifying deformations of  $\bar{r}$  which are minimal outside  $Q$ , in the sense of Definition 4.6. Choose any surjection  $R_\infty \rightarrow R_{Q_N}$  (possible by Proposition 4.10), and let  $R_\infty \rightarrow R$  be the composite of this surjection with the natural map  $R_{Q_N} \rightarrow R^{\min}$ .

We define the module  $H_N$  as follows. Let  $\Delta$  be the unique quotient of  $\Delta_{Q_N} = \prod_{x \in Q_N} (\mathbf{Z}/x)^\times$  which is isomorphic to  $(\mathbf{Z}/p^N \mathbf{Z})^q$ , and let  $X_\Delta(Q_N) \rightarrow X_0(Q_N)$  be as in Section 7.2. Let  $\mathfrak{m}_N$  be the ideal  $\mathfrak{m} \subset \mathbf{T}_\sigma(Q_N)$  of Theorem 7.2 when  $Q$  is taken to be  $Q_N$ . We then take

$$H_N := H^0(X_\Delta(Q_N), \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha, \beta, \vee}_{\mathfrak{m}_N}$$

and we regard it as an  $R_\infty$ -module via the surjection  $R_\infty \rightarrow R_{Q_N}$  chosen above, and the classifying map  $R_{Q_N} \rightarrow \mathbf{T}_\sigma(Q)_{\mathfrak{m}_N}^{\alpha, \beta}$  associated to the deformation  $r_{Q_N}$  of Theorem 6.17. The  $S_N$ -module structure on  $H_N$  is given by choosing an identification  $\Delta \cong (\mathbf{Z}/p^N \mathbf{Z})^q$ .

We need to check that, given these definitions, the conditions of Proposition 3.3 hold.

- (a) The image of  $S_N$  in  $\text{End}_{\mathcal{O}}(H_N)$  is contained in the image of  $R_\infty$  because under the Galois representation  $r_{Q_N}$  of Theorem 6.17, the image of an element  $\sigma \in I_x$ , for  $x$  a prime in  $Q_N$ , is conjugate to a matrix of the form  $\text{diag}(1, 1, \langle u \rangle, \langle u \rangle)$  where  $\text{Art}_x(u) = \sigma$ . This follows from [64, Corollary 3].
- (b) We have

$$\begin{aligned} (H_N)_{\Delta_N} &= \left( H^0(X_\Delta(Q_N), \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha, \beta}_{\mathfrak{m}_N} \right)^{\Delta_N \vee} \\ &= H^0(X_0(Q_N), \omega(a, 2)(-\infty)_{K/\mathcal{O}})^{\alpha, \beta, \vee}_{\mathfrak{m}_N}. \end{aligned}$$

Combining this with the isomorphism of Theorem 7.2, we obtain an isomorphism:

$$\psi_N : (H_N)_{\Delta_N} \xrightarrow{\sim} H.$$

(c) Finally,  $H_N$  is finite and balanced over  $S_N$  by Theorem 7.11. We can thus apply Proposition 3.3, and we deduce that  $H$  is a finite free  $R$ -module. Since the action of  $R$  on  $H$  factors through  $\mathbf{T}_{\sigma, m\emptyset}^{\alpha, \beta}$ , the conclusions of Theorem 9.1 follow immediately.  $\square$

## Appendix. Bloch–Kato conjectures for automorphic motives

FRANK CALEGARI, DAVID GERAGHTY, and MICHAEL HARRIS

### A.1. Introduction

Assume (for this paragraph only) the standard conjectures, and suppose that  $M$  is a pure irreducible Grothendieck motive over  $\mathbf{Q}$  with coefficients in (say) a totally real field  $E$ . We make no assumption on the regularity or self-duality of  $M$ . According to conjectures of Hasse–Weil, Langlands, Clozel, and others, one expects that the motive  $M$  is automorphic, and corresponds to an algebraic cuspidal automorphic representation  $\pi$  for  $\mathrm{GL}(n)/\mathbf{Q}$  such that  $L(\pi, s) = L(M, s)$ . By a result of Jacquet and Shalika [43, Proposition 3.6], the  $L$ -function

$$L(M \times M^\vee, s) = L(\pi \times \pi^\vee, s)$$

is meromorphic for  $\mathrm{Re}(s) > 0$  and has a simple pole at  $s = 1$ . Let  $\mathrm{ad}^0(M)$  be the pure motive of weight 0 with coefficients in  $E$  such that  $\mathrm{ad}^0(M) \oplus E = M \times M^\vee$ . Then

$$L(\mathrm{ad}^0(M), s) = \frac{L(\pi \times \pi^\vee, s)}{\zeta(s)},$$

and  $L(\mathrm{ad}^0(M), 1) \neq 0$  is finite. According to conjectures of Deligne and Bloch–Kato [10], for any pure de Rham representation  $V$ , there is an equality

$$\dim H_f^1(G_{\mathbf{Q}}, V) - \dim H^0(G_{\mathbf{Q}}, V) = \mathrm{ord}_{s=1} L(V^*, s).$$

In particular, if we take  $V = V^* = \mathrm{ad}^0(M)$ , then we expect that  $H_f^1(G_{\mathbf{Q}}, \mathrm{ad}^0(M))$  should vanish. This is a special case of the more general fact that  $H_f^1(\mathbf{Q}, V)$  should be trivial for any  $p$ -adic representation  $V$  arising from a pure motive  $M$  of weight  $w \geq 0$ . One also conjectures that the value of  $L(\mathrm{ad}^0(M), 1)$ , after normalization by some suitable period should lie in  $\mathbf{Q}^\times$ . Moreover, after equating  $M$  with its étale realization for some prime  $p$ , the normalized  $L$ -function should have the same valuation as the order of a corresponding Selmer group  $H_f^1(\mathbf{Q}, \mathrm{ad}^0(M) \otimes \mathbf{Q}_p/\mathbf{Z}_p)$ .

No longer assuming any conjectures, suppose that  $M = \{r_\lambda\}$  is now a weakly compatible system of  $n$ -dimensional irreducible Galois representations of  $G_{\mathbf{Q}}$ , and suppose moreover that  $M$  is automorphic, that is, it corresponds to a cuspidal form  $\pi$  for  $\mathrm{GL}(n)/\mathbf{Q}$  in a manner compatible with the local Langlands correspondence.



Then, even without the standard conjectures, it makes sense to ask, for a  $p$ -adic representation  $r : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_n(\mathcal{O})$  coming from  $M$  (for some finite extension  $K/\mathbf{Q}_p$  with ring of integers  $\mathcal{O}$ ), if the Selmer group

$$H_f^1(\mathbf{Q}, \mathrm{ad}^0(r) \otimes K/\mathcal{O})$$

is finite. Theorems of this kind were first proved for  $n = 2$  by Flach [26], and they are also closely related to modularity lifting theorems as proved by Taylor and Wiles [66] and Wiles [73] (see in particular [24]). More precisely, the order of this group is related to the order of a congruence ideal between modular forms. In this paper, we prove versions of these results for modular abelian surfaces and (conditionally) compatible families of  $n$ -dimensional representations whose existence was only recently proved to exist in [38]. The main theorem is the following.

**THEOREM A.1**

Let  $A/\mathbf{Q}$  be a semistable modular abelian surface with  $\mathrm{End}(A) = \mathbf{Z}$ . Let  $p$  be a prime such that

- (1)  $p$  is sufficiently large with respect to some constant depending only on  $A$ ;
- (2)  $A$  is ordinary at  $p$ , and if  $\alpha, \beta$  are the unit root eigenvalues of  $D_{\mathrm{cris}}(V)$ , then

$$(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2\beta^2 - 1) \not\equiv 0 \pmod{p}.$$

Then

$$H_f^1(\mathbf{Q}, \mathrm{asp}^0(r) \otimes \mathbf{Q}_p/\mathbf{Z}_p) = 0,$$

where  $\mathrm{asp}^0(r)$  is the 10-dimensional adjoint representation of  $\mathrm{PGSp}(4)$ . Moreover, the set of primes  $p$  satisfying these conditions has density 1.

For the families of Galois representations constructed in [38], we prove the following result.

**THEOREM A.2**

Let  $\pi$  be a weight 0 regular algebraic cuspidal representation for  $\mathrm{GL}(n)/F$  for a CM field  $F$  and coefficients in  $E$ . Let  $\lambda$  be a prime of  $\mathcal{O}_E$  dividing  $p$ , and let

$$r = r_{\lambda}(\pi) : G_F \rightarrow \mathrm{GL}_n(\mathcal{O})$$

be a  $p$ -adic representation associated to  $\pi$  with determinant  $\epsilon^{n(1-n)/2}$ . Assume the following.

- (1) The representation  $\bar{r}|_{F(\xi_p)}$  has enormous image in the sense of [16, Section 9.2].

- (2) *Let  $v \neq p$  be a prime at which  $\pi$  is ramified.*
- (a)  *$\pi_v$  is an unramified twist of the Steinberg representation.*
  - (b) *The representations  $r|_{G_{I_v}}$  and  $\bar{r}|_{G_{I_v}}$  are unipotent. For a topological generator  $\sigma_v \in I_v$  of tame inertia,  $\bar{r}(\sigma)$  consists of a single block, namely:*

$$\dim \ker(\bar{r}(\sigma) - \text{id}_n) = 1.$$

- (3) *We have that  $p$  is sufficiently large with respect to some constant depending only on  $\pi$ .*

*Assume all of [16, Conjecture B] except assumption (4). Then the Selmer group  $H_f^1(F, \text{ad}^0(r) \otimes K/\mathcal{O})$  is trivial.*

Note that [16, Conjecture B] consists of five parts. The first part concerns local-global compatibility at  $v \mid p$ , which is still open. The second and third parts concern local-global compatibility at finite  $v$ . Here there is work in characteristic 0 by Varma [71], although arguments of this nature should also apply to the Galois representations constructed by Scholze [60], at least for modularity lifting purposes (since for modularity lifting it is usually sufficient to have local-global compatibility up to  $N$ -semisimplification). The fifth part is essentially addressed in [16], and also (in a different and arguably superior manner) in [46]. Hence the main remaining issue is local-global compatibility at  $\ell = p$ .

Unlike the case of Theorem A.1, we do not know whether Theorem A.2 applies for infinitely many  $p$ . One reason is that we do not even know that the representations  $r_\lambda$  are irreducible for sufficiently large  $p$ . Another is that we do not know whether  $r$  is a minimal deformation of  $\bar{r}$  at ramified primes  $v$  for sufficiently large  $p$ , although this is predicted to hold by some generalization of Serre's conjecture. One example to which this does apply is to the Galois representations associated to symmetric powers of non-CM elliptic curves  $E$  over  $F$ . (The conclusion of the theorem holds for  $F$  if it holds for any extension  $F'/F$ , and any symmetric power of  $E/F$  is potentially modular over some CM extension  $F'/F$  in this case by [1].) We deduce our theorems from the modularity lifting results of [16], for which we assume familiarity. One obstruction to directly applying the theorems of [16] is that the modularity results of [16] require further unproved assumptions, namely, the vanishing of certain cohomology groups outside a prescribed range. The main observation here is that vanishing in these cases may be established for all sufficiently large  $p$ .

For automorphic representations for  $\text{GL}(n)$ , we require the extra assumption of local-global compatibility at  $v \mid p$ , which is not yet known in full generality. Some results along this line have very recently been announced in [1], although they are not strong enough to give a completely unconditional proof of Theorem A.2. One problem is that [1, Theorem 4.5.1] requires the hypothesis that  $F$  contains an imaginary

quadratic field in which  $p$  splits, which has to fail for a set of primes  $p$  of positive density. It may be possible to give an unconditional version of Theorem A.2 under some such assumption on  $p$ , although we do not pursue this here, in part because we would still be unable to establish for a general  $\pi$  that condition (b) holds for infinitely many  $p$ . (The nilpotent ideal of [60] and [1] would also be an annoying complication.) Similarly, the assumption (a) that  $\pi_v$  is a twist of the Steinberg representation can (in principle) be weakened to the weaker assumption that  $\pi_v$  is of the form  $\mathrm{Sp}_{n_1}(\chi_1) \boxplus \mathrm{Sp}_{n_2}(\chi_2) \cdots \boxplus \mathrm{Sp}_{n_k}(\chi_k)$  for some partition  $n = \sum_{i=1}^k n_i$ . The main reason we do *not* do this is that it would require a more precise discussion of local-global compatibility at  $v \nmid p$ , and in particular a refinement of [16, Conjecture B].

### A.2. Relation with special values of periods

The Bloch–Kato conjecture actually gives a more precise prediction of the exact order of the Selmer group in terms of the value of the  $L$ -function divided by a certain motivic period. One can think of this as two separate conjectures. The first is to show that the normalization of  $L(1)$  by a suitable period is indeed rational. The second is to relate the corresponding  $p$ -adic valuation of this ratio to the order of a Selmer group. Our method naturally relates the order of a Selmer group to a certain tangent space. On the other hand, for most of the Galois representations we consider, it is not known whether there exists a corresponding motive, and so it is not clear exactly what it means to prove rationality. There are some formulations where one can establish certain forms of rationality (or even integrality) with respect to periods defined in terms of automorphic integrals (see, e.g., [5], [34], and also [69]). However, it is not clear to the authors how these results exactly relate to the (sometimes conjectural) motivic periods. An interesting test case is the following. Suppose that

$$\rho : G_{\mathbf{Q}} \rightarrow \mathrm{GL}_2(\mathbf{C})$$

is an irreducible odd representation. According to the Artin conjecture (known in this case; see [14], [15], [47]–[49]), one knows that  $\rho$  is modular of weight 1. If one chooses a prime  $p$ , and supposes that  $\rho$  has a model over  $\mathcal{O}$ , then the finiteness of the Selmer group  $H_f^1(\mathbf{Q}, \mathrm{ad}^0(\rho) \otimes K/\mathcal{O})$  is a consequence of the finiteness of the  $p$ -class group of  $\mathbf{Q}(\ker(\rho))$ . (The former is a quotient of the latter.) The methods of this paper (following [16]) show that, at all primes  $p > 2$  such that  $\rho$  is unramified, the Selmer group  $H_f^1(\mathbf{Q}, \mathrm{ad}^0(\rho) \otimes K/\mathcal{O})$  is detected by congruences between the modular form  $f$  and other Katz modular forms of weight 1 which may not lift to characteristic 0. In particular, there exist such congruences if and only if  $H_f^1(\mathbf{Q}, \mathrm{ad}^0(\rho) \otimes K/\mathcal{O})$  is nonzero. However, unlike in the case of higher weight modular forms, there does not seem to be an a priori way to relate this to a normalization of the adjoint  $L$ -function  $L(\mathrm{ad}^0(\rho), 1)$  (which in this case is an Artin  $L$ -function). The issue is that all such constructions (following Hida [42]) proceed by understanding various pairings on the

Betti cohomology of arithmetic groups in characteristic 0, whereas weight 1 Katz modular forms only have an interpretation in terms of coherent cohomology. Even in cases where one does have access to Betti cohomology, say for regular algebraic cuspidal automorphic representations for  $\mathrm{GL}(n)/F$  (even for  $\mathrm{GL}(2)$  over imaginary quadratic fields  $F$ ), it is not so clear whether the cohomological pairings one can define give the “correct” regulators or merely the regulators up to some finite multiple related to the torsion classes in cohomology. Since we have nothing to say about how to resolve these issues, we follow Wittgenstein’s dictum in [74, Section 7] and say no more about them.

### A.3. Vanishing theorems

The main idea of this present article is to note that the various vanishing theorems which are required inputs for the method of [16] may be established at least for  $p$  sufficiently large. This is not so useful for applications to modularity—if  $p$  is sufficiently large, then any completion of the appropriate Hecke ring  $\mathbf{T}$  at a maximal ideal  $\mathfrak{m}$  of residue characteristic  $p$  will be formally smooth of dimension 1, and so the only characteristic 0 representation that one can prove is modular is the representation one must assume is modular in the first place. However, with respect to Selmer groups, this statement does have content—it says that these representations will have no infinitesimal deformations.

#### A.3.1. Betti cohomology

Let  $F$  be an imaginary CM field of degree  $2d$ . Let

$$\begin{aligned} l_0 &:= d(\mathrm{rank}(\mathrm{SL}_n(\mathbf{C})) - \mathrm{rank}(\mathrm{SU}_n(\mathbf{C}))) = d(n - 1), \\ 2q_0 + l_0 &= d(\mathrm{dim}(\mathrm{SL}_n(\mathbf{C})) - \mathrm{dim}(\mathrm{SU}_n(\mathbf{C}))) = d(n^2 - 1), \\ q_0 &= \frac{d(n^2 - n)}{2}. \end{aligned}$$

Fix a tempered cuspidal automorphic representation  $\pi$  for  $\mathrm{PGL}(n)/F$  of weight 0 with coefficients in  $E$ . Let  $Y = Y(K)$  be the corresponding arithmetic orbifold considered in [16, Section 9], where  $K$  is chosen to be maximal at all unramified primes for  $\pi$  and Iwahori level  $\mathrm{Iw}_v$  for all ramified primes. Let  $\mathbf{T}$  denote the (anemic) Hecke algebra defined as the  $\mathbf{Z}$ -subalgebra of

$$\mathrm{End} \bigoplus_{k,m} H^k(Y(K), \mathbf{Z}/m\mathbf{Z})$$

generated by Hecke endomorphisms  $T_{\alpha,i}$  for  $i \leq n$  and  $\alpha$  which are units at primes dividing the level (cf. [16, Definition 9.1]). For a prime  $v$  of  $\mathcal{O}_E$ , let

$$\bar{r}_v : G_F \rightarrow \mathrm{GL}_n(k)$$

be the corresponding semisimple Galois representation, and let  $\mathfrak{m}$  denote the corresponding maximal ideal of  $\mathbf{T}$ .

LEMMA A.3

*For all sufficiently large  $v$ , and  $\mathcal{O} = \mathcal{O}_{E,v}$ , we have  $H^i(Y, \mathcal{O}/\varpi^k)_{\mathfrak{m}} = 0$  unless  $i \in [q_0, \dots, q_0 + l_0]$ .*

*Proof*

Assume otherwise. Pick a neat finite index subgroup  $K' \subset K$ , and a corresponding Galois cover  $Y' = Y(K') \rightarrow Y = Y(K)$  where  $Y'$  is now a manifold. It follows that  $H^*(Y', \mathbf{Z})$  is finitely generated, and thus  $H^*(Y, \mathbf{Z}[1/M])$  is also finitely generated where  $M$  denotes the product of primes dividing  $[K : K']$ . We now assume that  $v$  has residue characteristic prime to  $[K : K']$ . Since  $H^*(Y, \mathbf{Z}[1/M])$  is finitely generated, the groups  $H^*(Y, \mathcal{O}) = H^*(Y, \mathbf{Z}[1/M]) \otimes \mathcal{O}$  are torsion-free and of finite rank over  $\mathcal{O}$  for all  $i$  when  $\mathcal{O}$  has sufficiently large residue characteristic. Moreover, there exist only finitely many systems of eigenvalues which occur in  $H^*(Y, \mathbf{R})$ . Assuming that the result is false (and there are infinitely many  $v$ ), we deduce that there exists an eigenclass  $[c]$  in  $H^i(Y, \mathcal{O}_E)$  with  $i \notin [q_0, \dots, q_0 + l_0]$  such that the action of  $\mathbf{T}$  on  $[c]$  has support at  $\mathfrak{m}$  for infinitely many primes  $v$  of  $\mathcal{O}_E$ . By the Chinese remainder theorem, the Hecke eigenvalues of  $[c]$  coincide with those of  $\pi$ . We now show that  $[c]$  corresponds to an automorphic form  $\Pi$  which must simultaneously be nontempered and yet isomorphic to  $\pi$ , giving a contradiction. Eigenclasses in cohomology may be realized by isobaric automorphic representations (see [28, Theorem 2.3]). Suppose that  $[c]$  corresponds to such an automorphic representation  $\Pi$ . Because of the degree where  $[c]$  occurs, we deduce (from [11, Chapter II, Proposition 3.1] and [19, Lemme 3.14]) that  $\Pi$  is not tempered. Yet by strong multiplicity 1 (see [43]), there is an isomorphism  $\Pi \simeq \pi$ .  $\square$

For a more detailed discussion (in a more general setting) relating the cohomology of local systems to tempered automorphic representations, see the proof of [1, Theorem 2.4.9].

THEOREM A.4

*Suppose that  $H^i(Y, \mathcal{O}/\varpi^n)_{\mathfrak{m}} = 0$  unless  $i \in [q_0, \dots, q_0 + l_0]$ . Let  $Q$  be a finite collection of primes  $x$  such that  $\bar{r}(\text{Frob}_x)$  has distinct eigenvalues and  $N(x) \equiv 1 \pmod{p}$ . Then*

$$H^i(Y_1(Q), \mathcal{O}/\varpi^n)_{\mathfrak{m}_\alpha} = 0$$

*for all  $i \notin [q_0, \dots, q_0 + l_0]$ , where*

- (1)  $\alpha = \{\alpha_x\}$  is a choice of eigenvalues of  $\bar{r}(\text{Frob}_x)$  for each  $x$  dividing  $Q$ ;
- (2) the localization takes place with respect to the Hecke algebra  $\mathbf{T}_Q$  consisting of the Hecke operators prime to  $p$  and prime to the level together with  $U_x - \alpha_x$  for all  $x \in Q$ .

In particular, the conclusions of this theorem apply for all sufficiently large  $p$ .

*Proof*

We first note that the assumption (of absolute irreducibility) on  $\bar{r}$  ensures that the cohomology of the boundary vanishes after localization at  $\mathfrak{m}$ . This is because the cohomology of the boundary may be computed inductively from the cohomology of Levi subgroups and then of  $\text{GL}_{n_i}$  for  $n_i < n$  (see, e.g., Section 3 and in particular Proposition 3.3 of [17]), and so the corresponding Galois representations associated to these classes are reducible.

By Poincaré duality (and the discussion above concerning the vanishing of the boundary cohomology localized at  $\mathfrak{m}$ ), it suffices to prove the result for  $i < q_0$ . Let  $i$  be the smallest integer for which the inequality is violated. Then, by the Hochschild–Serre spectral sequence, we deduce that

$$H^i(Y_0(Q), k)_{\mathfrak{m}_\alpha} \neq 0.$$

As in [16, Section 9.4] (see also [46, Lemma 6.25(4)]), we deduce that  $H^i(Y_0(Q), k)_{\mathfrak{m}_\alpha} \simeq H^i(Y, k)_{\mathfrak{m}}$ . The result then follows by Lemma A.3.  $\square$

The modularity lifting theorems of [16] are proved by constructing sets of so-called *Taylor–Wiles primes* which have the property that imposing local conditions at these primes annihilates (as much as possible) the dual Selmer group. The assumption that  $\bar{r}_v|_{F(\xi_p)}$  has enormous image implies that there exist arbitrarily many sets  $Q$  of auxiliary Taylor–Wiles primes satisfying the hypothesis that  $\bar{r}(\text{Frob}_x)$  has distinct eigenvalues. In particular, Theorem A.4 serves as a replacement for Conjecture B(4) of [16]. (For a different (and somewhat cleaner) treatment of Taylor–Wiles primes using the enormous image hypothesis, see [46], which is also used in [1].)

*A.3.2. Coherent cohomology*

Let  $\mathcal{O}$  denote the ring of integers in some finite extension of  $\mathbf{Q}_p$ . Let  $X$  denote a toroidal compactification of a Siegel 3-fold  $Y$  of level prime to  $p$  over  $\text{Spec } \mathcal{O}$ , and let  $Z$  denote the minimal compactification. Let  $\pi : \mathcal{A} \rightarrow Y$  denote the universal abelian variety, let  $\mathcal{E} = \pi_* \Omega_{\mathcal{A}/X}^1$ , let  $\omega = \det \mathcal{E}$ , and, by abuse of notation, also let  $\omega$  denote the canonical extension of  $\omega$  to  $X$  or the corresponding ample line bundle on  $Z$ . Fix a cuspidal automorphic representation  $\pi$  for  $\text{GSp}(4)/\mathbf{Q}$  corresponding to a modular abelian surface  $A$  which we assume has endomorphism ring  $\mathbf{Z}$  over  $\mathbf{Q}$ , and hence to

a cuspidal Siegel modular form of scalar weight 2. Let

$$\bar{r}_p : G_{\mathbf{Q}} \rightarrow \mathrm{GSp}_4(\mathbf{F}_p)$$

be the corresponding semisimple representation for each prime  $p$ . Let  $\mathfrak{m}$  denote the corresponding maximal ideal of  $\mathbf{T}$ .

LEMMA A.5

*For all sufficiently large  $p$ , and any set  $Q$  of auxiliary primes, we have*

$$H^i(X, \omega_{\mathcal{O}/\mathfrak{w}^n}^2)_{\mathfrak{m}_Q} = 0$$

*for  $i = 2$  and  $3$ , where  $\mathfrak{m}_Q$  denotes the maximal ideal in the Hecke algebra where operators at  $Q$  and  $p$  have been omitted.*

*Proof*

This is a consequence of the proof of Theorem 7.11 above. We give a brief sketch here of the idea: as in the proof of Lemma A.3, we otherwise deduce that there exists a characteristic 0 form in  $H^i(X, \omega_{\mathcal{O}}^2)$  giving rise to infinitely many of these classes. The representation  $\bar{r}_p$  will be irreducible for all sufficiently large  $p$  (because  $\mathrm{End}_{\mathbf{Q}}(A) = \mathbf{Z}$ ; see also the proof of Lemma A.7). It follows that the transfer of this form to  $\mathrm{GL}(4)$  must be cuspidal and, moreover (by multiplicity 1), coincide with the transfer of the representation coming from the holomorphic Siegel modular form. But such a representation only contributes to cohomology in degrees 0 and 1.  $\square$

LEMMA A.6

*Suppose that  $H^2(X, \omega_{\mathcal{O}/\mathfrak{w}^n}^2)_{\mathfrak{m}_Q} = 0$  for any set of auxiliary primes  $Q$ , as in the statement of Lemma A.5. Suppose, moreover, that  $\bar{r}_p$  is absolutely irreducible. Then, for  $i \geq 2$ ,*

$$H^i(X_1(Q), \omega_{\mathcal{O}/\mathfrak{w}^n}^2)_{\mathfrak{m}_\alpha} = H^i(X_0(Q), \omega_{\mathcal{O}/\mathfrak{w}^n}^2)_{\mathfrak{m}_\alpha} = 0,$$

*where  $Q$  is any collection of primes where  $\bar{r}(\mathrm{Frob}_x)$  has distinct eigenvalues,  $N(x) \equiv 1 \pmod{p}$ , and  $\alpha = \{\alpha_x\}$  is any collection of eigenvalues of  $\rho(\mathrm{Frob}_x)$  for  $x$  dividing  $Q$ . In particular, the conclusions of this theorem apply for all sufficiently large  $p$ .*

*Proof*

A somewhat elaborate version of this result is proved in Lemmas 7.4 and 7.5 above. Here, however, we use instead the modified treatment of Taylor–Wiles primes by Khare and Thorne (cf. [46, Lemma 6.25]), as adapted for  $\mathrm{GSp}(4)$  in [13, Sections 2.4 and 7.9], which leads to a great simplification. By dévissage, we can reduce to the case where the coefficients are a finite field  $k$ . To deduce vanishing for  $X_1(Q)$ , we

first use Serre duality to reduce the problem to vanishing of  $H^i(X_1(Q), \omega(-\infty))_{\mathfrak{m}_\alpha^*}$  for  $i \leq 1$ . (Serre duality is only Hecke equivariant up to a twist by a power of the cyclotomic character, which is recorded by the star.) Equivalently, it suffices to show that, if  $i$  denotes the smallest integer such that  $H^i(X_1(Q), \omega(-\infty))_{\mathfrak{m}_\alpha^*}$  is nonzero, then  $H^i(X, \omega(-\infty))_{\mathfrak{m}_\mathcal{O}}$  is also nonzero. By Hochschild–Serre applied to the map  $X_1(Q) \rightarrow X_0(Q)$ , it suffices to show that, if  $i$  denotes the smallest integer such that  $H^i(X_0(Q), \omega(-\infty))_{\mathfrak{m}_\alpha^*}$  is nonzero, then  $H^i(X, \omega(-\infty))_{\mathfrak{m}_\mathcal{O}^*}$  is also nonzero. The result now follows as in the proof of [13, Proposition 7.9.8], which identifies these groups for all  $i$  under the Taylor–Wiles hypothesis using [13, Lemmas 2.4.36 and 2.4.37].  $\square$

#### A.4. Proofs

Let  $R$  denote the minimal deformation ring of  $\bar{\rho}$  defined as follows:

- (1) *Coherent case:*  $R$  is the minimal ordinary deformation ring denoted by  $R^{\min} = R_\emptyset$  in [16, Section 4].
- (2) *Betti case:*  $R$  is the minimal ordinary deformation ring corresponding to the following conditions:
  - (a) If  $v$  is a prime of bad reduction (so we are assuming, for a topological generator  $\sigma$  of tame inertia, that  $\bar{\rho}(\sigma)$  is unipotent with a single block), then we take the local deformation ring to be the ring  $R_v^1$  in [16, Section 8.5.1]. Note that, if  $r(\sigma)$  on  $A^n$  has characteristic polynomial  $(X - 1)^n$ , then (given our assumption on  $\bar{\rho}$ ) this deformation problem coincides with the minimal condition in [20, Definition 2.4.14], namely, that the map

$$\ker(r(\sigma) - \text{id}_n)^r \otimes_R k \rightarrow \ker(\bar{\rho}(\sigma) - \text{id}_n)^r$$

is an isomorphism (equivalently, surjection) for all  $r \leq n$ . (One can see this equivalence by induction— $\bar{\rho}(\sigma)$  has a unique eigenvector over  $k$ , which lifts to a unique eigenvector over  $A$  whose mod- $p$  reduction is nontrivial; now take the representation of  $A^{n-1}$  and  $k^{n-1}$  given by quotienting out by this eigenvector.)

- (b) If  $v \mid p$  and  $p$  is sufficiently large with respect to  $n$  and the primes which ramify in  $F$ , then we take deformations which are Fontaine–Laffaille of weight  $[0, 1, \dots, n - 1]$ .

If one has an isomorphism  $R \simeq \mathbf{T}$  for all sufficiently large  $p$  satisfying the required hypothesis, then since one also will have an isomorphism  $\mathbf{T} \simeq \mathcal{O}$ , this would immediately imply that the tangent space to  $R$  along the projection to  $\mathcal{O}$  is trivial, and hence the corresponding adjoint Selmer groups are trivial. Theorem A.2 is now a consequence of [16, Theorems 5.16 and 6.4], where we use the fact that the



corresponding local deformation rings are formally smooth (as follows from [20, Corollary 2.4.3, Lemma 2.4.19]), and where we use Theorem A.4 as a substitute for the vanishing assumption required in [16, Conjecture B]. Equally, Theorem A.1 follows as in Theorem 1.2 above, where the vanishing result of Lemma A.6 replaces the vanishing results of Lan and Suh [53] for other low-weight local systems. The modularity argument above requires a large image hypothesis which we assume in Theorem A.2 and which we are required to prove (for sufficiently large  $p$ ) for Theorem A.4. Furthermore, we must justify the claim in Theorem A.4 that the assumptions hold for a set of primes  $p$  of density 1. Hence it remains to prove the following.

LEMMA A.7

Let  $A$  be a semistable abelian surface of conductor  $N$  with  $\text{End}(A) = \mathbf{Z}$ . Then

- (1) for sufficiently large  $p$ , the residual representation  $\bar{r}_p : G_{\mathbf{Q}} \rightarrow \text{GSp}_4(\mathbf{F}_p)$  is surjective with minimal conductor  $N$ ;
- (2) for a set of primes  $p$  of density 1, we have

$$\alpha\beta(\alpha^2 - 1)(\beta^2 - 1)(\alpha - \beta)(\alpha^2\beta^2 - 1) \not\equiv 0 \pmod{p}.$$

*Proof*

Since  $\text{End}(A) = \mathbf{Z}$ , the residual image is surjective for all sufficiently large  $p$  by [62, Corollaire au Théorème 3]. In order to ensure that the conductor of  $\bar{r}_p$  at a prime  $\ell$  dividing  $N$  matches that of  $A$ , it suffices to take  $p$  co-prime to the (finite) order of the component group  $\Phi_{\mathcal{A}}$  of the Néron model of  $A^{\vee}$  at  $\ell$ . This proves the first claim.

For the second claim, let the characteristic polynomial of Frobenius (for  $p$  not dividing the discriminant on the étale cohomology  $V_{\ell} = H^1(A, \overline{\mathbf{Q}}_{\ell})$  at any prime  $\ell \neq p$ ) be

$$X^4 + a_p X^3 + (2p + b_p)X^2 + pa_p X + p^2.$$

Let the roots of this polynomial be  $\alpha$ ,  $\beta$ ,  $\alpha^{-1}p$  and  $\beta^{-1}p$  respectively; by the Riemann hypothesis for curves (Weil bound) they are Weil numbers of absolute value  $\sqrt{p}$ . Note that  $2p + b_p$  is the trace of  $\text{Frob}_p$  on  $\wedge^2 V_{\ell}$  for all but finitely many  $\ell$ , and  $a_p^2$  is the trace of  $\text{Frob}_p$  on  $V_{\ell} \otimes V_{\ell}$ . We use the following lemma, which is essentially an observation of Deligne, Milne, Ogus, and Shih [22, Proposition 2.7(1)].

LEMMA A.8

There is no fixed linear relation between  $1$ ,  $p$ ,  $b_p$ ,  $a_p$ , and  $a_p^2$  which can hold for a set  $p \in S$  of positive density.

*Proof*

From such an equality, we can build two finite-dimensional representations  $A_\ell$  and  $B_\ell$  built out of copies of  $\wedge^2 V_\ell$ ,  $\mathbf{Q}_p$ ,  $\mathbf{Q}_p(1)$ ,  $V_\ell$ , and  $V_\ell \otimes V_\ell$ , respectively, which have equal trace on  $\text{Frob}_p$  for infinitely many  $p$ . There must be at least one quadratic field with a positive density of inert primes in  $S$ ; twisting by this representation we arrive at a representation  $W_\ell$  with a set  $S$  of positive density such that  $\text{Frob}_p$  has trace zero for  $p \in S$ . For sufficiently large  $\ell$ , our assumptions on  $A$  imply (see [62]) that the image of  $G_{\mathbf{Q}}$  on  $V_\ell$  is  $\text{GSp}_4(\mathbf{Z}_\ell)$  if  $\text{End}(A) = \mathbf{Z}$ . By Chebotarev, it follows that the appropriate identity must also hold on a subset of these groups of positive measure. Yet the distribution of the appropriate eigenvalues for  $\text{GSp}_4(\mathbf{Z}_\ell)$  does not have any atomic measure. In particular, writing the eigenvalues in either case as  $x$ ,  $y$ ,  $\delta/x$ , and  $\delta/y$ , we would obtain a relation between the polynomials

$$\begin{aligned} 1, \quad \delta, \quad xy + \delta x/y + \delta y/x + \delta^2/xy, \\ x + y + \delta/x + \delta/y, \quad (x + y + \delta/x + \delta/y)^2 \end{aligned}$$

that holds on an open set (and consequently holds everywhere). There are no such relations by inspection.  $\square$

Returning to the proof of Lemma A.7, we deduce from the Weil bounds that  $|a_p|^2 \leq 16p$  and  $|b_p| \leq 4p$ . Hence, if we have any linear expression in  $a_p$ ,  $a_p^2$ ,  $b_p$  and 1 which is congruent to zero modulo  $p$  for a set of positive density, then it must also equal a constant multiple of  $p$  for a set of positive density, and we would obtain a contradiction by Lemma A.8. We show that this holds in each of the possible cases when our congruence above holds. (We take advantage of the symmetry in  $\alpha$  and  $\beta$  and consider a reduced number of cases.)

- (1) Suppose that neither  $\beta$  nor  $\beta^{-1}p$  are units. Then  $b_p \equiv 0 \pmod{p}$ .
- (2) Suppose that  $\alpha\beta \equiv \varepsilon \pmod{p}$  for some fixed  $\varepsilon \in \{\pm 1\}$ . Then  $b_p \equiv \varepsilon \pmod{p}$ .
- (3) Suppose that  $\alpha = \varepsilon \pmod{p}$  for some fixed  $\varepsilon \in \{\pm 1\}$ . Then  $b_p - \varepsilon a_p + 1 \equiv 0 \pmod{p}$ .
- (4) Suppose that  $\alpha - \beta \equiv 0 \pmod{p}$ . Then  $4b_p - a_p^2 \equiv 0 \pmod{p}$ .  $\square$

*Acknowledgments.* We would like to thank George Boxer for some very helpful comments related to the proofs of Theorems 8.10 and 8.11 and for pointing out an oversight in the first version of the Appendix. We would also like to thank Olivier Taïbi for answering some technical questions arising in Section 7.2. We would also like to acknowledge useful conversations with Kevin Buzzard, Ching-Li Chai, Matthew Emerton, Toby Gee, Michael Harris, Kai-Wen Lan, Vincent Pilloni, and Jack Thorne. The participants of the 2014 Bellairs workshop in number theory, where an earlier version of this paper was discussed, deserve our thanks and recognition. And finally,

we thank the referees, whose detailed comments very much helped to improve this manuscript.

Calegari's work was partially supported by National Science Foundation (NSF) Career grant DMS-0846285 and by NSF grants DMS-1404620 and DMS-1701703. Geraghty's work was partially supported by NSF grants DMS-1200304 and DMS-1128155. Harris's work was partially supported by NSF grant DMS-1404769. Harris also received support from the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC "Arithmetic of automorphic motives (AAMOT)" grant 290766.

## References

- [1] P. B. ALLEN, F. CALEGARI, A. CARAIANI, T. GEE, D. HELM, B. V. LE HUNG, J. NEWTON, P. SCHOLZE, R. TAYLOR, and J. A. THORNE, *Potential automorphy over CM fields*, preprint, [arXiv:1812.09999v1](https://arxiv.org/abs/1812.09999v1) [math.NT]. (82, 83, 85, 86)
- [2] A. N. ANDRIANOV, *Quadratic Forms and Hecke Operators*, Grundlehren Math. Wiss. **286**, Springer, Berlin, 1987. [MR 0884891](https://doi.org/10.1007/978-3-642-70341-6). [DOI 10.1007/978-3-642-70341-6](https://doi.org/10.1007/978-3-642-70341-6). (58)
- [3] J. ARTHUR, "Automorphic representations of  $\mathrm{GSp}(4)$ " in *Contributions to Automorphic Forms, Geometry, and Number Theory*, Johns Hopkins Univ. Press, Baltimore, MD, 2004, 65–81. [MR 2058604](https://doi.org/10.1090/coll/061). (6, 40, 54, 55)
- [4] ———, *The Endoscopic Classification of Representations: Orthogonal and Symplectic Groups*, Amer. Math. Soc. Colloq. Publ. **61**, Amer. Math. Soc., Providence, 2013. [MR 3135650](https://doi.org/10.1090/coll/061). [DOI 10.1090/coll/061](https://doi.org/10.1090/coll/061). (6)
- [5] B. BALASUBRAMANYAM and A. RAGHURAM, *Special values of adjoint  $L$ -functions and congruences for automorphic forms on  $\mathrm{GL}(n)$  over a number field*, Amer. J. Math. **139** (2017), no. 3, 641–679. [MR 3650229](https://doi.org/10.2307/2501707). [DOI 10.1353/ajm.2017.0017](https://doi.org/10.2307/2501707). (83)
- [6] J. BELLAÏCHE and G. CHENEVIER, *Families of Galois representations and Selmer groups*, Astérisque **324**, Soc. Math. France, Paris 2009. [MR 2656025](https://doi.org/10.1017/S0010437X11005264). (39)
- [7] ———, *The sign of Galois representations attached to automorphic forms for unitary groups*, Compos. Math. **147** (2011), no. 5, 1337–1352. [MR 2834723](https://doi.org/10.1112/S0010437X11005264). [DOI 10.1112/S0010437X11005264](https://doi.org/10.1112/S0010437X11005264). (40)
- [8] J. BERGSTRÖM, C. FABER, and G. VAN DER GEER, *Siegel modular forms of genus 2 and level 2: Cohomological computations and conjectures*, Int. Math. Res. Not. IMRN **2008**, art. ID 100. [MR 2439544](https://doi.org/10.1093/imrn/rnn100). [DOI 10.1093/imrn/rnn100](https://doi.org/10.1093/imrn/rnn100). (31)
- [9] D. BLASIUS, M. HARRIS, and D. RAMAKRISHNAN, *Coherent cohomology, limits of discrete series, and Galois conjugation*, Duke Math. J. **73** (1994), no. 3, 647–685. [MR 1262930](https://doi.org/10.1215/S0012-7094-94-07326-2). [DOI 10.1215/S0012-7094-94-07326-2](https://doi.org/10.1215/S0012-7094-94-07326-2). (10, 24, 26, 28, 29)
- [10] S. BLOCH and K. KATO, "L-functions and Tamagawa numbers of motives" in *The Grothendieck Festschrift, Vol. I*, Progr. Math. **86**, Birkhäuser Boston, Boston, 1990, 333–400. [MR 1086888](https://doi.org/10.1007/BF02924888). (80)

- [11] A. BOREL and N. R. WALLACH, *Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups*, Ann. of Math. Stud. **94**, Princeton Univ. Press, Princeton, 1980. [MR 0554917](#). (85)
- [12] G. BOXER, *Torsion in the coherent cohomology of Shimura varieties and Galois representations*, Ph.D. dissertation, Harvard University, Cambridge, MA, 2015. [MR 3450451](#). (49)
- [13] G. BOXER, F. CALEGARI, T. GEE, and V. PILLONI, *Abelian surfaces over totally real fields are potentially modular*, preprint, [arXiv:1812.09269v1](#) [math.NT]. (4, 5, 71, 87, 88)
- [14] K. BUZZARD, *Analytic continuation of overconvergent eigenforms*, J. Amer. Math. Soc. **16** (2003), no. 1, 29–55. [MR 1937198](#). [DOI 10.1090/S0894-0347-02-00405-8](#). (3, 83)
- [15] K. BUZZARD and R. TAYLOR, *Companion forms and weight one forms*, Ann. of Math. (2) **149** (1999), no. 3, 905–919. [MR 1709306](#). [DOI 10.2307/121076](#). (3, 83)
- [16] F. CALEGARI and D. GERAGHTY, *Modularity lifting beyond the Taylor-Wiles method*, Invent. Math. **211** (2018), no. 1, 297–433. [MR 3742760](#). [DOI 10.1007/s00222-017-0749-x](#). (3, 4, 6, 11, 12, 53, 67, 71, 81, 82, 83, 84, 86, 88, 89)
- [17] A. CARAIANI and B. V. LE HUNG, *On the image of complex conjugation in certain Galois representations*, Compos. Math. **152** (2016), no. 7, 1476–1488. [MR 3530448](#). [DOI 10.1112/S0010437X16007016](#). (86)
- [18] F. CLÉRY and G. VAN DER GEER, *Constructing vector-valued Siegel modular forms from scalar-valued Siegel modular forms*, Pure Appl. Math. Q. **11** (2015), no. 1, 21–47. [MR 3394973](#). [DOI 10.4310/PAMQ.2015.v11.n1.a2](#). (58)
- [19] L. CLOZEL, “Motifs et formes automorphes: applications du principe de fonctorialité” in *Automorphic Forms, Shimura Varieties, and L-functions, Vol. I (Ann Arbor, 1988)*, Perspect. Math. **10**, Academic Press, Boston, 1990, 77–159. [MR 1044819](#). (55, 85)
- [20] L. CLOZEL, M. HARRIS, and R. TAYLOR, *Automorphy for some  $l$ -adic lifts of automorphic mod  $l$  Galois representations*, with Appendix B “Unipotent representations of  $\mathrm{GL}(n, F)$  in the quasi-banal case” by M.-F. Vignéras, Publ. Math. Inst. Hautes Études Sci. **108** (2008), 1–181. [MR 2470687](#). [DOI 10.1007/s10240-008-0016-1](#). (40, 43, 88, 89)
- [21] H. DARMON, F. DIAMOND, and R. TAYLOR, “Fermat’s last theorem” in *Elliptic Curves, Modular Forms and Fermat’s Last Theorem (Hong Kong, 1993)*, Int. Press, Cambridge, MA, 1997, 2–140. [MR 1605752](#). (16, 17)
- [22] P. DELIGNE, J. S. MILNE, A. OGUS, and K. SHIH, *Hodge Cycles, Motives, and Shimura Varieties*, Lecture Notes in Math. **900**, Springer, Berlin, 1982. [MR 0654325](#). (89)
- [23] P. DELIGNE and J.-P. SERRE, *Formes modulaires de poids 1*, Ann. Sci. Éc. Norm. Supér. (4) **7** (1974), 507–530 (1975). [MR 0379379](#). (31)
- [24] F. DIAMOND, M. FLACH, and L. GUO, *The Tamagawa number conjecture of adjoint motives of modular forms*, Ann. Sci. Éc. Norm. Supér. (4) **37** (2004), no. 5, 663–727. [MR 2103471](#). [DOI 10.1016/j.ansens.2004.09.001](#). (81)

- [25] G. FALTINGS, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366. MR 0718935. DOI 10.1007/BF01388432. (20)
- [26] M. FLACH, *A finiteness theorem for the symmetric square of an elliptic curve*, Invent. Math. **109** (1992), no. 2, 307–327. MR 1172693. DOI 10.1007/BF01232029. (81)
- [27] M. FLANDER, *A theta operator for Siegel modular forms (mod  $p$ )*, Ph.D. dissertation, University of Melbourne, Melbourne, 2013. (61)
- [28] J. FRANKE and J. SCHWERMER, *A decomposition of spaces of automorphic forms, and the Eisenstein cohomology of arithmetic groups*, Math. Ann. **311** (1998), no. 4, 765–790. MR 1637980. DOI 10.1007/s002080050208. (85)
- [29] T. GEE and D. GERAGHTY, *Companion forms for unitary and symplectic groups*, Duke Math. J. **161** (2012), no. 2, 247–303. MR 2876931. DOI 10.1215/00127094-1507376. (43)
- [30] T. GEE and O. TAÏBI, *Arthur’s multiplicity formula for  $\mathrm{GSp}_4$  and restriction to  $\mathrm{Sp}_4$* , J. Éc. Polytech. Math. **6** (2019), 469–535. MR 3991897. (6)
- [31] A. GENESTIER and J. TILOUINE, “Systèmes de Taylor-Wiles pour  $\mathrm{GSp}(4)$ ” in *Formes automorphes, II: Le cas du groupe  $\mathrm{GSp}(4)$* , Astérisque **302**, Soc. Math. France, Paris, 2005, 177–290. MR 2234862. (4, 17, 44, 48)
- [32] W. GOLDRING, *Galois representations associated to holomorphic limits of discrete series*, Compos. Math. **150** (2014), no. 2, 191–228. MR 3177267. DOI 10.1112/S0010437X13007355. (37)
- [33] W. GOLDRING and J.-S. KOSKIVIRTA, *Strata Hasse invariants, Hecke algebras and Galois representations*, Invent. Math. **217** (2019), no. 3, 137–151. MR 3989256. DOI 10.1007/s00222-019-00882-5. (49)
- [34] H. GROBNER, M. HARRIS, and E. LAPID, “Whittaker rational structures and special values of the Asai  $L$ -function” in *Advances in the Theory of Automorphic Forms and Their  $L$ -functions*, Contemp. Math. **664**, Amer. Math. Soc., Providence, 2016, 119–134. MR 3502979. DOI 10.1090/conm/664/13108. (83)
- [35] B. H. GROSS, *A tameness criterion for Galois representations associated to modular forms (mod  $p$ )*, Duke Math. J. **61** (1990), no. 2, 445–517. MR 1074305. DOI 10.1215/S0012-7094-90-06119-8. (60)
- [36] A. GROTHENDIECK, M. RAYNAUD, and D. S. RIM, *Groupes de monodromie en géométrie algébrique*, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 I), Lecture Notes in Math. **288**, Springer, Berlin, 1972. Zbl 0237.00013. MR 0354656. (31)
- [37] M. HARRIS, “Automorphic forms and the cohomology of vector bundles on Shimura varieties” in *Automorphic Forms, Shimura Varieties, and  $L$ -functions, Vol. II (Ann Arbor, 1988)*, Perspect. Math. **11**, Academic Press, Boston, 1990, 41–91. MR 1044828. (25, 28, 55)
- [38] M. HARRIS, K.-W. LAN, R. TAYLOR, and J. THORNE, *Galois representations for regular algebraic cusp forms over CM-fields*, in preparation. (81)
- [39] M. HARRIS and S. ZUCKER, *Boundary cohomology of Shimura varieties, I: Coherent cohomology on toroidal compactifications*, Ann. Sci. Éc. Norm. Supér. (4) **27** (1994), no. 3, 249–344. MR 1272293. (53)

- [40] ———, *Boundary cohomology of Shimura varieties, III: Coherent cohomology on higher-rank boundary strata and applications to Hodge theory*, *Mém. Soc. Math. Fr. (N.S.)* **2001**, no. 85. [MR 1850830](#). [DOI 10.24033/msmf.398](#). (52, 53)
- [41] F. HERZIG and J. TILOUINE, *Conjecture de type de Serre et formes compagnons pour  $\mathrm{GSp}_4$* , *J. Reine Angew. Math.* **676** (2013), 1–32. [MR 3028753](#). [DOI 10.1515/CRELLE.2011.190](#). (8)
- [42] H. HIDA, *Congruence of cusp forms and special values of their zeta functions*, *Invent. Math.* **63** (1981), no. 2, 225–261. [MR 0610538](#). [DOI 10.1007/BF01393877](#). (83)
- [43] H. JACQUET and J. A. SHALIKA, *On Euler products and the classification of automorphic forms, II*, *Amer. J. Math.* **103** (1981), no. 4, 777–815. [MR 0623137](#). [DOI 10.2307/2374050](#). (80, 85)
- [44] P. L. KASSAEI, *A gluing lemma and overconvergent modular forms*, *Duke Math. J.* **132** (2006), no. 3, 509–529. [MR 2219265](#). [DOI 10.1215/S0012-7094-06-13234-9](#). (3, 4)
- [45] N. M. KATZ, “A result on modular forms in characteristic  $p$ ” in *Modular Functions of One Variable, V (Bonn, 1976)*, *Lecture Notes in Math.* **601**, Springer, Berlin, 1977, 53–61. [MR 0463169](#). (66)
- [46] C. KHARE and J. A. THORNE, *Potential automorphy and the Leopoldt conjecture*, *Amer. J. Math.* **139** (2017), no. 5, 1205–1273. [MR 3702498](#). [DOI 10.1353/ajm.2017.0030](#). (4, 82, 86, 87)
- [47] C. KHARE and J.-P. WINTENBERGER, *Serre’s modularity conjecture, I*, *Invent. Math.* **178** (2009), no. 3, 485–504. [MR 2551763](#). [DOI 10.1007/s00222-009-0205-7](#). (83)
- [48] ———, *Serre’s modularity conjecture, II*, *Invent. Math.* **178** (2009), no. 3, 505–586. [MR 2551764](#). [DOI 10.1007/s00222-009-0206-6](#). (83)
- [49] M. KISIN, *Modularity of 2-adic Barsotti-Tate representations*, *Invent. Math.* **178** (2009), no. 3, 587–634. [MR 2551765](#). [DOI 10.1007/s00222-009-0207-5](#). (83)
- [50] N. KOBLITZ,  *$p$ -adic variation of the zeta-function over families of varieties defined over finite fields*, *Compos. Math.* **31** (1975), no. 2, 119–218. [MR 0414557](#). (62)
- [51] K.-W. LAN, *Arithmetic Compactifications of PEL-Type Shimura Varieties*, *London Math. Soc. Monogr. Ser.* **36**, Princeton Univ. Press, Princeton, 2013. [MR 3186092](#). [DOI 10.1515/9781400846016](#). (21)
- [52] K.-W. LAN and J. SUH, *Vanishing theorems for torsion automorphic sheaves on compact PEL-type Shimura varieties*, *Duke Math. J.* **161** (2012), no. 6, 1113–1170. [MR 2913102](#). [DOI 10.1215/00127094-1548452](#). (22, 23, 27)
- [53] ———, *Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties*, *Adv. Math.* **242** (2013), 228–286. [MR 3055995](#). [DOI 10.1016/j.aim.2013.04.004](#). (22, 23, 24, 89)
- [54] J. S. MILNE, “Canonical models of (mixed) Shimura varieties and automorphic vector bundles” in *Automorphic Forms, Shimura Varieties, and  $L$ -functions, Vol. I (Ann Arbor, 1988)*, *Perspect. Math.* **10**, Academic Press, Boston, 1990, 283–414. [MR 1044823](#). (27)

- [55] C. P. MOK, *Galois representations attached to automorphic forms on  $GL_2$  over CM fields*, *Compos. Math.* **150** (2014), no. 4, 523–567. MR 3200667. DOI 10.1112/S0010437X13007665. (39, 55)
- [56] V. PILLONI, *Modularité, formes de Siegel et surfaces abéliennes*, *J. Reine Angew. Math.* **666** (2012), 35–82. MR 2920881. DOI 10.1515/CRELLE.2011.123. (4, 13, 17, 18, 60)
- [57] ———, *Sur la théorie de Hida pour le groupe  $GSp_{2g}$* , *Bull. Soc. Math. France* **140** (2012), no. 3, 335–400. MR 3059119. DOI 10.24033/bsmf.2630. (21, 22, 36)
- [58] ———, *Higher coherent cohomology and  $p$ -adic modular forms of singular weight*, preprint, available at <http://hal.archives-ouvertes.fr/hal-01393374v3/document> (accessed 15 January 2020). (4, 5)
- [59] V. PILLONI and B. STROH, *Surconvergence et classicit : le cas Hilbert*, *J. Ramanujan Math. Soc.* **32** (2017), no. 4, 355–396. MR 3733761. (4)
- [60] P. SCHOLZE, *On torsion in the cohomology of locally symmetric varieties*, *Ann. of Math. (2)* **182** (2015), no. 3, 945–1066. MR 3418533. DOI 10.4007/annals.2015.182.3.3. (82, 83)
- [61] J.-P. SERRE, *Propri t s galoisiennes des points d’ordre fini des courbes elliptiques*, *Invent. Math.* **15** (1972), no. 4, 259–331. MR 0387283. DOI 10.1007/BF01405086. (20)
- [62] ———, “Lettre   Marie-France Vign ras du 10/2/1986” in *Collected Papers, IV: 1985–1998*, Springer Collect. Works Math., Springer, Berlin, 2000, 38–55. MR 1730973. (89, 90)
- [63] C. SKINNER and E. URBAN, *Sur les d formations  $p$ -adiques de certaines repr sentations automorphes*, *J. Inst. Math. Jussieu* **5** (2006), no. 4, 629–698. MR 2261226. DOI 10.1017/S147474800600003X. (33)
- [64] C. M. SORENSEN, *Galois representations attached to Hilbert-Siegel modular forms*, *Doc. Math.* **15** (2010), 623–670. MR 2735984. (29, 43, 44, 79)
- [65] R. TAYLOR, *Galois representations associated to Siegel modular forms of low weight*, *Duke Math. J.* **63** (1991), no. 2, 281–332. MR 1115109. DOI 10.1215/S0012-7094-91-06312-X. (38)
- [66] R. TAYLOR and A. WILES, *Ring-theoretic properties of certain Hecke algebras*, *Ann. of Math. (2)* **141** (1995), no. 3, 553–572. MR 1333036. DOI 10.2307/2118560. (81)
- [67] J. TILOUINE, *Siegel varieties and  $p$ -adic Siegel modular forms*, *Doc. Math.* **2006**, no. Extra Vol., 781–817. MR 2290605. (56, 57)
- [68] ———, *Formes compagnons et complexe BGG dual pour  $GSp_4$* , *Ann. Inst. Fourier (Grenoble)* **62** (2012), no. 4, 1383–1436. MR 3025747. (33, 36)
- [69] E. URBAN, *Module de congruences pour  $GL(2)$  d’un corps imaginaire quadratique et th orie d’Iwasawa d’un corps CM biquadratique*, *Duke Math. J.* **92** (1998), no. 1, 179–220. MR 1611003. DOI 10.1215/S0012-7094-98-09204-3. (83)
- [70] G. VAN DER GEER, “Cycles on the moduli space of abelian varieties” in *Moduli of Curves and Abelian Varieties*, Aspects Math. **E33**, Vieweg, Braunschweig, 1999, 65–89. MR 1722539. DOI 10.1007/978-3-322-90172-9\_4. (62)
- [71] I. VARMA, *Local-global compatibility for regular algebraic cuspidal automorphic representations when  $\ell \neq p$* , preprint, arXiv:1411.2520v1 [math.NT]. (82)

- [72] G. WIESE, *On Galois representations of weight one*, Doc. Math. **19** (2014), 689–707.  
[MR 3247800](#). (71)
- [73] A. WILES, *Modular elliptic curves and Fermat’s last theorem*, Ann. of Math. (2) **141**  
(1995), no. 3, 443–551. [MR 1333035](#). [DOI 10.2307/2118559](#). (81)
- [74] L. WITTGENSTEIN, *Tractatus Logico-Philosophicus*, Harcourt, Brace and Company,  
New York, 1922. (84)
- [75] T. YAMAUCHI, *The weight reduction of mod  $p$  Siegel modular forms for  $\mathrm{GSp}_4$* ,  
preprint, [arXiv:1410.7894v5](#) [math.NT]. (61, 62)

*Calegari*

Department of Mathematics, University of Chicago, Chicago, Illinois, USA;  
[fcale@math.uchicago.edu](mailto:fcale@math.uchicago.edu)

*Geraghty*

Junior Ganymede Club, London, UK; [djgeraghty@gmail.com](mailto:djgeraghty@gmail.com)

*Harris*

Department of Mathematics, University of Chicago, Chicago, Illinois, USA;  
[harris@math.columbia.edu](mailto:harris@math.columbia.edu)