ABELIAN SURFACES OVER TOTALLY REAL FIELDS ARE POTENTIALLY MODULAR

GEORGE BOXER, FRANK CALEGARI, TOBY GEE, AND VINCENT PILLONI

Abstract. We show that abelian surfaces (and consequently curves of genus 2) over totally real fields are potentially modular. As a consequence, we obtain the expected meromorphic continuation and functional equations of their Hasse–Weil zeta functions. We furthermore show the modularity of infinitely many abelian surfaces A over $\mathbf{Q}$ with $\mathrm{End}_{\mathbf{C}} A = \mathbf{Z}$. We also deduce modularity and potential modularity results for genus one curves over (not necessarily CM) quadratic extensions of totally real fields.

Contents

1. Introduction 3
1.1. Our main theorems 3
1.2. An overview of our argument 6
1.3. An outline of the paper 12
1.4. Some further remarks 13
1.5. Acknowledgements 15
2. Background material 15
2.1. Notation and conventions 15
2.2. Induction of two-dimensional representations 17
2.3. The non-archimedean local Langlands correspondence 18
2.4. Local representation theory 19
2.5. Purity 31
2.6. Archimedean L-parameters 32
2.7. Galois representations associated to automorphic representations 33
2.8. Compatible systems of Galois representations, L-functions, and Hasse–Weil zeta functions 35
2.9. Arthur's classification 38
2.10. Balanced modules 39
2.11. Projectors 40
3. Shimura varieties 41
3.1. Similitude groups 41
3.2. Shimura varieties over $\mathbf{C}$ 42
3.3. Integral models of Shimura varieties 43
3.4. Local models 49
3.5. Compactifications 50
3.6. Functorialities 53
3.7. Automorphic vector bundles 55
3.8. Coherent cohomology and Hecke operators 56
3.9. Cohomological correspondences — definitions 62
3.10. Cohomology and automorphic representations 70

G.B. was supported in part by NSF postdoctoral fellowship DMS-1503047. F.C. was supported in part by NSF Grants DMS-1404620, DMS-1701703, and DMS-2001097. T.G. was supported in part by a Leverhulme Prize, EPSRC grant EP/L025485/1, ERC Starting Grant 306326, and a Royal Society Wolfson Research Merit Award. V.P was supported in part by the ANR-14-CE25-0002-01 Percolator, and the ERC-2018-COG-818856-HiCoShiVa.

4. Hida complexes 73
4.1. Mod p-geometry: Hasse invariants and stratifications 75
4.2. Vanishing theorem for ordinary cohomology 77
4.3. Formal geometry 95
4.4. Sheaves of p-adic modular forms for G_1 98
4.5. Hecke operators at p on the cohomology of p-adic modular forms 99
4.6. Perfect complexes of p-adic modular forms 105
5. Doubling 116
5.1. The case of $\mathrm{GL}_2/\mathbf{Q}$ 117
5.2. Conventions 119
5.3. The operator $U_{\mathrm{Kli}(w),1}$ 119
5.4. The operators $U_{\mathrm{Iw}(w),1}$ and Z_w 121
5.5. Preliminaries on Kodaira–Spencer 122
5.6. The Hecke operator $U_{\mathrm{Iw}(w),1}$ and traces for partial Frobenius 124
5.7. The Hecke operator $U_{w,2}$ on the w-non ordinary locus. 128
5.8. Main doubling results 132
5.9. Vanishing in partial negative weight: Fourier–Jacobi expansions 135
6. Higher Coleman theory 138
6.1. Sheaves of overconvergent and locally analytic modular forms: the formal construction 138
6.2. Vanishing theorem 143
6.3. Sheaves of overconvergent and locally analytic modular forms: the analytic construction 146
6.4. Locally analytic overconvergent classes and algebraic overconvergent classes 153
6.5. Small slope forms are classical 154
6.6. Application to ordinary cohomology 164
7. The Taylor–Wiles/Calegari–Geraghty method 167
7.1. Galois deformation rings 169
7.2. Galois cohomology and presentations 171
7.3. Local deformation problems, $l = p$ 172
7.4. Local deformation problems, $l \neq p$ 188
7.5. Big image conditions and vast representations 198
7.6. Taylor–Wiles primes 208
7.7. Global Galois deformation problems 210
7.8. Taylor–Wiles systems: initial construction 211
7.9. Taylor–Wiles systems: local-global compatibility 216
7.10. An abstract patching criterion 223
7.11. The patching construction 224
7.12. Cycles and modules over products of local deformation rings 226
7.13. Multiplicities of patched spaces of p-adic automorphic forms 229
8. Étale descent and the main modularity lifting theorem 235
8.1. Introduction 235
8.2. Étale descent 237
8.3. Solvable base change 246
8.4. The main modularity lifting theorem 247
8.5. Base change and automorphy lifting 248
9. Potential modularity of abelian surfaces 249
9.1. Compatible systems and potential automorphy 249
9.2. Abelian surfaces 253
9.3. Potential modularity and meromorphic continuation 258
9.4. K3 surfaces of large rank 261
10. Applications to modularity 264
10.1. First modularity results 264
10.2. Abelian varieties with fixed 3-torsion 268
10.3. The Paramodular Conjecture 272
10.4. Examples and counterexamples 276
References 282

1. Introduction

1.1. **Our main theorems.** Let X be a smooth, projective variety of dimension m over a number field F with good reduction outside a finite set of primes S. Associated to X, one may write down a global Hasse–Weil zeta function:

$$\zeta_X(s) = \prod \frac{1}{1 - N(x)^{-s}},$$

where the product runs over all the closed points x of some (any) smooth proper integral model $\mathcal{X}/\mathcal{O}_F[1/S]$ for X. (We suppress S from the notation — different choices of S only change $\zeta_X(s)$ by a finite number of Euler factors.) The function $\zeta_X(s)$ is absolutely convergent for $\mathrm{Re}(s) > 1 + m$. We have the following:

Conjecture 1.1.1 (Hasse–Weil Conjecture, cf. [Ser70], in particular Conj. C9)**.** *The function $\zeta_X(s)$ extends to a meromorphic function of $\mathbf{C}$. There exists a positive real number $A \in \mathbf{R}^{>0}$, non-zero rational functions $P_v(T)$ for $v|S$, and infinite Gamma factors $\Gamma_v(s)$ for $v|\infty$ such that:*

$$\xi(s) = \zeta_X(s) \cdot A^{s/2} \cdot \prod_{v|\infty} \Gamma_v(s) \cdot \prod_{v|S} P_v(N(v)^{-s})$$

satisfies the functional equation $\xi(s) = w \cdot \xi(m + 1 - s)$ with $w = \pm 1$.

(In Serre's formulation of the conjecture, the Gamma factors are also given explicitly in terms of the Archimedean Hodge structures of X.) This conjecture appears to be first formulated in print (albeit in a less precise form and only for curves) on the final page of [Wei52]. If $F = \mathbf{Q}$ and X is a point, then $\zeta_X(s)$ is the Riemann zeta function, and Conjecture 1.1.1 follows from Riemann's functional equation [Rie59]. If F is a general number field but X is still a point, then $\zeta_X(s)$ is the Dedekind zeta function $\zeta_F(s)$, and Conjecture 1.1.1 is a theorem of Hecke [Hec20]. If X is a curve of genus zero, then (up to bad Euler factors) $\zeta_X(s) = \zeta_F(s)\zeta_F(s-1)$, and Conjecture 1.1.1 follows immediately. More generally, if X is any smooth projective variety whose cohomology is generated by algebraic cycles over $\overline{F}$, then $\zeta_X(s)$ is a finite product of Artin L-functions (up to translation), and Conjecture 1.1.1 in this case is a consequence of Brauer's theorem [Bra47]. In the case when the Galois representations associated to the l-adic cohomology of X are *potentially* abelian (e.g. an abelian variety with CM), Conjecture 1.1.1 is also a consequence of the results of Hecke and Brauer.

The fundamental work of Wiles [Wil95, TW95] and the subsequent work of Breuil, Conrad, Diamond, and Taylor [CDT99, BCDT01] proved Conjecture 1.1.1 for curves $X/\mathbf{Q}$ of genus one, since (again up to a finite number of Euler factors) $\zeta_X(s) = \zeta_{\mathbf{Q}}(s)\zeta_{\mathbf{Q}}(s-1)/L(E,s)$ (where $E = \mathrm{Jac}(X)$), and the modularity of E implies the holomorphy and functional equation for $L(E,s)$. More generally, the potential modularity results of [Tay02] imply Conjecture 1.1.1 for curves X/F of genus one over any totally real field. The methods used in these papers have been vastly generalized over the past 25 years due to the enormous efforts of many people. On the other hand, these methods have until recently been extremely reliant on the assumption that the Hodge numbers $h^{p,q} = \dim H^{p,q}_{\mathrm{dR}}(X) = \dim H^p(X, \Omega^q)$ of X are at most 1 for all p and q, or at least that such an inequality holds (suitably interpreted) for the irreducible motives occurring in the cohomology of X. While many such motives exist inside the cohomology of Shimura varieties, there is a paucity of natural geometric examples satisfying this condition. For example, if X

is a curve of genus g, then $h^{1,0} = h^{0,1} = g$, and so the original Taylor–Wiles method only applies when $g = 0$ or 1. For genus two curves, we prove the following theorem.

Theorem 1.1.2. *Let X be either a genus two curve or an abelian surface over a totally real field F. Then Conjecture 1.1.1 holds for X.*

We prove Theorem 1.1.2 as a corollary of the following theorem.

Theorem 1.1.3. *Let X be either a genus two curve or an abelian surface over a totally real field F. Then X is potentially automorphic.*

Here by *potentially automorphic* we mean that there exists a finite Galois extension L/F such that the compatible system of Galois representations $\mathcal{R}$ attached to $H^1(X_{\overline{\mathbf{Q}}}, \mathbf{Q}_p)$ (as p varies) over L is automorphic in a precisely circumscribed sense which we make explicit in Definition 9.1.1. (See also Remark 9.1.9 for a discussion of how we distinguish between *automorphic* and *modular* in this paper; this distinction is made purely for technical convenience, and can safely be ignored while reading this introduction.) In particular, an immediate consequence is that the L-function of $H^1(X_{\overline{\mathbf{Q}}}, \mathbf{Q}_p)$ *as a G_L-representation* extends to a holomorphic function on all of $\mathbf{C}$. Theorem 1.1.2 follows from Theorem 1.1.3 via a standard argument with Brauer's theorem and base change, together (in the case of abelian surfaces) with known functorialities in small rank. (Some care must be taken in this deduction if the p-adic Galois representations associated to X become reducible after restriction to L; this issue does not arise in the most interesting cases of Theorem 1.1.3, in particular the case of an abelian surface X with $\mathrm{End}_{\mathbf{C}}(X) = \mathbf{Z}$.)

Theorem 1.1.3 (and thus also Theorem 1.1.2) is a consequence of Theorem 9.3.1 and Corollary 9.3.3, which in turn are deduced from our main modularity lifting theorem, Theorem 8.4.1. As a consequence of Theorem 1.1.3, we also deduce the following potential modularity result for genus one curves (see Theorem 9.3.4):

Theorem 1.1.4. *Let X be a genus one curve over a quadratic extension K/F of a totally real field F. Then X is potentially modular.*

When K/F is totally real, this result has been known for some time ([Tay02]). When K/F is totally imaginary, however, the result was only recently proved in [ACC+18]. For all other quadratic extensions (such as $F = \mathbf{Q}(\sqrt{2})$ and $K = \mathbf{Q}(\sqrt[4]{2})$), the result is new. (See the remarks in §1.4.4 for a comparison between the methods of this paper with those of [ACC+18].)

Just as elliptic curves over $\mathbf{Q}$ can be associated (via the modularity theorem) to modular forms of weight 2, the Langlands program predicts that abelian surfaces over $\mathbf{Q}$ should be modular in the sense that they correspond to certain weight 2 Siegel modular forms. This is because (due to the existence of polarizations) the Galois representations associated to the p-adic Tate modules of abelian surfaces are naturally valued in $\mathrm{GSp}_4(\mathbf{Q}_p)$, and GSp_4 is its own Langlands dual group. A consideration of the Hodge–Tate weights then suggests that the corresponding automorphic forms on GSp_4 should be of weight 2 (see §10.3 for a more detailed discussion of this).

Our methods also have implications for the *modularity* (as opposed to potential modularity) of abelian surfaces over totally real fields. Here is an example of what can be proven by our methods.

Theorem 1.1.5. *There exist infinitely many modular abelian surfaces $A/\mathbf{Q}$ up to twist with* $\mathrm{End}_{\mathbf{C}} A = \mathbf{Z}$.

As a consequence, one deduces that the L-function associated to A in Theorem 1.1.5 (that is, the L-function associated to the Galois representation $H^1(A_{\overline{\mathbf{Q}}}, \mathbf{Q}_p)$ for any prime p) has a holomorphic continuation to the entire complex plane. Note that Theorem 1.1.3 only implies that this L-function has a meromorphic continuation, with no control over any possible poles. (This is for essentially the same reason that Brauer's theorem proves the meromorphic continuation of Artin L-functions, but not the holomorphic continuation.) In fact, we can also prove an analogous theorem for any totally real field F in which 3 splits completely; see Theorem 10.2.6.

To put Theorem 1.1.5 into context, note firstly that if $\mathrm{End}_{\mathbf{C}}(A) \neq \mathbf{Z}$, then the Galois representations associated to A become reducible over some finite extension, and hence one may use (or prove) special cases of functoriality to reduce the problem to the modularity of representations of dimensions 2 or 1. Results of this kind appear in the papers [Yos80, Yos84, RS07a, JLR12, DK16, BDPcS15]. (Several of these arguments could now be redone more systematically in light of the monumental work of Arthur [Art04, Art13].)

In the "typical" case that $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$, Brumer and Kramer [BK14] formulated the *paramodular conjecture*, which gives a precise prescription for the "optimal" level structure for an automorphic form corresponding to a given abelian surface; in particular, this in principle reduces the conjecture for a given A to an explicit computation of a (finite-dimensional) space of Siegel modular forms. They furthermore showed that the smallest prime conductor of an abelian surface is 277; in combination with the computations of [PY15], this demonstrates that the conjecture is true in prime conductor less than 277 (because there are neither any abelian surfaces nor suitable Siegel modular forms).

These considerations are taken further in the recent papers [BPP$^+$19, BK20]. In particular, these papers succeed in establishing for the first time the modularity of (finitely many, up to twist) abelian surfaces A with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$. (The explicit examples in [BPP$^+$19] are conductors 277, 353, and 587, and the example in [BK20] is of conductor 731. It should be noted that the abelian surfaces considered in Theorem 1.1.5 do not include any of these examples; as explained below, Theorem 1.1.5 is proved by proving the existence of infinitely many abelian surfaces to which our modularity lifting theorems apply, rather than by starting with explicit examples of small conductor.) These papers ultimately rely on elaborate explicit computations of low weight Siegel modular forms, developed in part by Poor and Yuen [PY15, PSY17, BPY16].

1.1.6. *Our modularity lifting theorem.* We now state our main modularity lifting theorem as it applies to abelian surfaces. The following theorem is proved in §10, see Proposition 10.1.1. (It is possible to slightly weaken the hypothesis at $v|p$ to deal with certain abelian surfaces which have semistable reduction at $v|p$.)

Theorem 1.1.7. *Let F be a totally real field in which $p > 2$ splits completely. Let A/F be an abelian surface with good ordinary reduction at all places $v|p$, and suppose that, at each $v|p$, the unit root crystalline eigenvalues are distinct modulo p. Assume that A admits a polarization of degree prime to p. Let*

$$\overline{\rho}_{A,p} : G_F \to \mathrm{GSp}_4(\mathbf{F}_p)$$

denote the dual of the mod-p Galois representation associated to $A[p]$, and assume that $\overline{\rho}_{A,p}$ is vast and tidy in the sense of Definitions 7.5.6 and 7.5.11. Assume

that $\overline{\rho}_{A,p}$ is ordinarily modular, in the sense that there exists an automorphic representation π of GSp_4/F of parallel weight 2 and central character $|\cdot|^2$ which is ordinary at all $v|p$, such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}_{A,p}$, and $\rho_{\pi,p}|_{G_{F_v}}$ is pure for all finite places v of F. Then A is modular, corresponding to a Hilbert–Siegel eigenform of parallel weight two.

Moreover, Proposition 10.1.3 shows that the modularity hypotheses on $\overline{\rho}_{A,p}$ can be omitted in the following situations:

(1) $p = 3$, and $\overline{\rho}_{A,3}$ is induced from a 2-dimensional representation with inverse cyclotomic determinant defined over a totally real quadratic extension E/F in which 3 is unramified.
(2) $p = 5$, and $\overline{\rho}_{A,5}$ is induced from a 2-dimensional representation valued in $\mathrm{GL}_2(\mathbf{F}_5)$ with inverse cyclotomic character defined over a totally real quadratic extension E/F in which 5 is unramified.
(3) $\overline{\rho}_{A,p}$ is induced from a character of a quartic CM field H/F in which p splits completely.

Theorem 1.1.7 may be viewed as the genus two analogue of [Wil95, Thm. 0.2], which is the main modularity lifting result proved in that paper. Proposition 10.1.3 is then the analogue of [Wil95, Thm. 0.6], which is a modularity result for residually projectively dihedral representations. The reason one cannot prove an analogue of [Wil95, Thm. 0.3] (which proves that all ordinary semistable elliptic curves over $\mathbf{Q}$ with $\overline{\rho}_{E,3}$ absolutely irreducible are modular) is that there is no argument to reduce the residual modularity of a surjective mod-3 representation $\overline{\rho}_3 : G_F \to \mathrm{GSp}_4(\mathbf{F}_3)$ (as in §5 of *ibid*) to special cases of the Artin Conjecture (proved by Langlands–Tunnell). Note that the difficulty is not simply that $\mathrm{GSp}_4(\mathbf{F}_3)$ is not solvable (some of the indicated representations above for $p = 3$ and 5 are non-solvable), but also that Artin representations do not contribute to the coherent cohomology of Shimura varieties in any setting other than holomorphic (Hilbert) modular forms of weight one.

For E/F a totally real quadratic extension, the inductions of (modular) representations $\overline{\varrho} : G_E \to \mathrm{GL}_2(\mathbf{F}_3)$ with determinant $\overline{\varepsilon}^{-1}$ to G_F provide a large source of residually modular $\overline{\rho}$. We then show that any such $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_3)$ with suitable determinant and local conditions at places $v|3$ is equal to $\overline{\rho}_{A,3}$ for infinitely many abelian surfaces A/F with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$ and with good ordinary reduction at $v|3$ (see Theorem 10.2.1). Theorem 1.1.7 then implies that all such A are modular, and hence implies Theorem 1.1.5.

1.2. **An overview of our argument.** Let A be an abelian surface over a totally real field F. We may assume that $\mathrm{End}_F(A) = \mathbf{Z}$ as otherwise, A is of GL_2-type, in which case it is known that A is potentially modular. If $\mathrm{End}_F(A) = \mathbf{Z}$, a generalization of the paramodular conjecture predicts the existence of a holomorphic weight 2 Hilbert–Siegel modular cuspidal eigenform f (for the group GSp_4/F) associated to A in the sense that we have an equality of L-functions $L(f, s) = L(H^1(A), s)$. If such an equality holds, we say that A is modular.

In this paper, we establish that (under some mild further restrictions on A), after possibly replacing the field F by a finite totally real extension F', the conjecture is true.

Remark 1.2.1. There are situations where we don't prove (even potentially) the paramodular conjecture for A. This is due to the presence of non-trivial endomorphisms of A over $\overline{\mathbf{Q}}$. Nevertheless, we always express the L-function of A using automorphic forms on groups GL_i/K for $i \in \{1, 2, 4\}$ and K a number field, and thus establish Conjecture 1.1.1.

On the surface, the modularity conjecture for abelian surfaces appears to be a generalization of the modularity conjecture for elliptic curves. However, this analogy is somewhat misleading. Elliptic curves are regular motives with weights $(0, 1)$, whereas abelian surfaces are irregular motives with weights $(0, 0, 1, 1)$. On the automorphic side, weight 2 Hilbert modular cuspforms occur in a single degree of the Betti and coherent cohomology of the Hilbert modular varieties. Under mild assumptions, there is an elliptic curve associated to any Hilbert modular cuspidal eigenform with rational Hecke field.

In contrast, weight 2 Hilbert–Siegel modular cuspforms only occur in the coherent cohomology of the Hilbert–Siegel modular variety. More precisely, a holomorphic weight 2 Hilbert–Siegel modular cuspidal eigenform can be viewed as a section of a line bundle ω^2 over the Hilbert–Siegel modular variety X; here X is a smooth algebraic variety defined over $\mathbf{Q}$ of dimension $3[F : \mathbf{Q}]$ which parametrizes abelian schemes of dimension $2[F : \mathbf{Q}]$ equipped with an action of $\mathcal{O}_F$, a level structure, and a polarization. Moreover, in the "generic case", such an eigenform contributes to cohomology in degrees 0 to $[F : \mathbf{Q}]$. Since the Hecke eigenvalues associated to such modular forms are not realized in the étale cohomology of a Shimura variety, we don't know how to associate a "motive" to a weight 2 Hilbert–Siegel modular cuspidal eigenform, but only a compatible system of Galois representations which should correspond to the system of ℓ-adic realizations of this motive. These Galois representations are constructed by using congruences.

From a technical point of view, it turns out that the modularity conjecture for abelian surfaces over a totally real field F is closely related to the 2-dimensional odd Artin conjecture for F (now a theorem), which is the existence of a bijection preserving L-functions between the following objects:

- Irreducible, totally odd, two dimensional complex representations of the absolute Galois group of F, and
- Hilbert modular cuspidal eigenforms (newforms) of weight one.

2-dimensional odd Artin representations have irregular Hodge–Tate weights $(0, 0)$, and Hilbert modular forms of weight one only occur in the coherent cohomology of the Hilbert modular variety, where they contribute in degrees 0 to $[F : \mathbf{Q}]$.

We now review some of the strategies employed in the proof of Artin's conjecture, as they have served as an inspiration for our current work. As with almost all modularity theorems, one proceeds by combining a modularity lifting theorem with residual modularity (that is, the modularity of the mod p representation). In the case of Artin's conjecture, residual modularity ultimately (if quite indirectly) comes from the Langlands–Tunnell theorem, whereas in our setting, the residual potential modularity comes from a straightforward application of Taylor's method [Tay02] using a theorem of Moret-Bailly. Accordingly, we ignore the question of residual modularity for the rest of this introduction, and concentrate on explaining the modularity lifting theorems.

The first modularity (lifting) theorems which applied to two dimensional odd Artin representations ρ over $\mathbf{Q}$ were obtained by Buzzard–Taylor and Buzzard [BT99,

Buz03]. There is an obstruction to generalizing the Taylor–Wiles method (which was originally applied in the regular case of Hodge–Tate weights $(0,1)$ and weight two modular forms [Wil95, TW95]) to the irregular case of weights $(0,0)$ and weight one modular forms. This obstruction lies in the fact that weight one forms occur in degrees 0 and 1 of the coherent cohomology and that there exist non-liftable mod p weight one eigenforms. (There is also a reflection of this obstruction on the Galois theoretic side — the corresponding local deformation ring at p has dimension one less in the irregular weight case.) Instead, Buzzard and Taylor proceed quite differently.

Choose a prime p and view ρ as a p-adic representation with finite image. We also assume that ρ is unramified at p and let α, β denote the Frobenius eigenvalues. For simplicity, we also assume that $\overline{\alpha} \neq \overline{\beta}$ (where the bar denotes reduction modulo p). We have that

$$\rho|_{G_{\mathbf{Q}_p}} \simeq \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \lambda_\beta \end{pmatrix}$$

for the unramified characters λ_α and λ_β taking a Frobenius element to α, β respectively.

The strategy of Buzzard and Taylor is to first replace the space of classical weight one modular forms by a bigger space of ordinary p-adic modular forms of weight one. On the Galois side, classical weight one eigenforms (of level prime to p) have associated Galois representations which are unramified at p, while an ordinary p-adic modular form f of weight one has an associated Galois representation which may be ramified at p of the form:

$$\rho_f|_{I_{\mathbf{Q}_p}} \simeq \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$$

Moreover, f should be classical if and only if $* = 0$. A key advantage of working with ordinary p-adic modular forms is that they are defined as sections of a line bundle over the ordinary locus, which is affine, and thus only occur in cohomological degree 0. It follows that ordinary p-adic modular forms of weight one are unobstructed for congruences and one can (assuming residual modularity) apply the Taylor–Wiles method in this setting to deduce the existence of two p-adic ordinary weight one modular forms f_α and f_β such that $\rho_{f_\alpha} = \rho_{f_\beta} = \rho$ and $U_p f_\alpha = \alpha f_\alpha$, $U_p f_\beta = \beta f_\beta$.

We observe that the existence of both f_α and f_β witnesses the fact that ρ is unramified at p. In order to show that f_α and f_β are classical forms of weight one, one forms the linear combinations $h = (\alpha f_\alpha - \beta f_\beta)/(\alpha - \beta)$ and $g = (f_\alpha - f_\beta)/(\alpha - \beta)$. The property that $\rho_{f_\alpha} = \rho_{f_\beta} = \rho$ and the explicit relation between q-expansions and Hecke eigenvalues translates into the geometric property that $\mathrm{Frob}(h) = g$. Using rigid analytic techniques, one can show that this property implies that f_α, f_β are classical forms of weight one. This strategy has been successfully generalized to any totally real field [Sas13, KST14, Kas16, PS16b, Pil17].

From a different direction, the paper [CG18] introduced an alternate method for proving modularity lifting results in weight one, by modifying the method of Taylor–Wiles and exploiting the Galois representations associated to coherent cohomology classes in all degrees. This method eliminates the delicate classicality theorem in weight one because one only works with classical (but possibly higher degree) cohomology. This method allows in principle to deal with any obstructed situation,

but requires some non-trivial input. For 2-dimensional odd Artin representations over a totally real fields, one needs to prove that (after suitable localization at a maximal ideal of the Hecke algebra) the cohomology in weight one is supported in degrees 0 to $[F : \mathbf{Q}]$ (this is actually automatic here for cohomological dimension reasons), and that the Galois representations in all cohomological degrees satisfy a form of local-global compatibility (at places above p). This last property has been proved when $F = \mathbf{Q}$ where one can reduce to studying degree 0 torsion cohomology classes and use the "doubling method" described below, but has not yet been proved for all primes p over a general totally real field (though see [ERX17] for some partial results).

After this discussion of Artin's conjecture, we return to the paramodular conjecture. We first assume that $F = \mathbf{Q}$ and fix a prime p. We assume that A has ordinary good reduction at p so that

$$\rho_{A,p}|_{G_{\mathbf{Q}_p}} \simeq \begin{pmatrix} \lambda_\alpha & 0 & * & * \\ 0 & \lambda_\beta & * & * \\ 0 & 0 & \lambda_\beta^{-1}\varepsilon^{-1} & 0 \\ 0 & 0 & 0 & \lambda_\alpha^{-1}\varepsilon^{-1} \end{pmatrix},$$

where, additionally, we assume that $\overline{\alpha} \neq \overline{\beta}$. (The Weil bounds together with the Cebotarev density theorem guarantee an ample source of such primes p.) Tilouine and his collaborators [TU95, Til98, TU99, MT02, GT05, Til06a, Til09] developed modularity lifting results for $\mathrm{GSp}_4/\mathbf{Q}$ in regular weight. In the case of Hodge–Tate weights $(0, 0, 1, 1)$, the paper [Pil12] applied these techniques to ordinary p-adic modular forms of weight 2 to produce (under technical assumptions) two p-adic eigenforms f_α and f_β associated to A (see also [Til06a, Til12], where the case of certain GSp_4-type abelian varieties is treated).

Similarly to the case of $\mathrm{GL}_2/\mathbf{Q}$, an ordinary p-adic modular form of weight 2 has a Galois representation whose restriction to inertia at p has the shape:

$$\begin{pmatrix} 1 & *_1 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & \varepsilon^{-1} & *_2 \\ 0 & 0 & 0 & \varepsilon^{-1} \end{pmatrix}.$$

Such a form should be classical if and only if its Galois representation is de Rham — equivalently: $*_1 = *_2 = 0$ (because of the symplectic structure, the vanishing of $*_1$ is equivalent to the vanishing of $*_2$).

As before, the existence of both f_α and f_β witnesses the property that A is de Rham at p. One difficulty, however, is that the Fourier expansions of Siegel modular forms are *not* explicitly determined by the Hecke eigenvalues (although we often have an abstract multiplicity one theorem). In particular, one doesn't know how to deduce geometrically from $\rho_{f_\alpha} = \rho_{f_\beta} = \rho_{A,p}$ that there exist suitable linear combinations of f_α and f_β giving rise to the desired form f by mimicking the Buzzard–Taylor argument.

In another direction, in [CG20] the modified Taylor–Wiles method was applied to low weight Siegel modular forms over $\mathbf{Q}$. There were a number of serious difficulties which prevented the authors from deducing any unconditional modularity lifting for abelian surfaces. The idea of the method is to consider (a suitable localization

of) the full cohomology complex $\mathrm{R}\Gamma(X, \omega^2)$ where X is an integral model over $\mathbf{Z}_p$ of the Siegel threefold. The required inputs are:

(1) to prove that the cohomology is only supported in degrees 0 and 1, and
(2) to prove local-global compatibility for the cohomology classes.

The first point is subtle in the weight of interest, because the cohomology groups will not generally vanish before localization at some non-Eisenstein maximal ideal $\mathfrak{m}$ (and indeed this point was not established in weight 2 in [CG20]). The paper [CG20] proved the second point for torsion degree 0 cohomology classes, using a "doubling" argument that we will return to below.

One crucial new ingredient which allows us to proceed in the symplectic case and deal with (1) is the higher Hida theory developed for GSp_4 over $\mathbf{Q}$ in [Pil20]. The idea of [Pil20] is (loosely speaking) to work over the larger space which is the complement of the supersingular locus (the rank ≥ 1 strata), which is now no longer affine. (Since we are working in mixed characteristic, one should imagine this taking place in the category of formal schemes, as in classical Hida theory.) Since the cohomological dimension of these spaces is one (more precisely, the image of these spaces in the minimal compactification has cohomological dimension one, which is sufficient for our purposes), there should exist complexes of amplitude $[0, 1]$ computing the coherent cohomology of all the relevant vector bundles. The main result of [Pil20] is that suitably constructed Hida idempotents cut down such a complex to a perfect complex, and moreover that the cohomology of this perfect complex is computed in characteristic zero by the space of weight 2 automorphic forms of interest. A crucial ingredient in order to study the coherent cohomology is therefore the introduction of Hecke operators at p and their associated projectors.

A version over $\mathbf{Q}$ of our modularity lifting theorem could be proved by applying the patching method of [CG18] to the higher Hida complexes of [Pil20]. It should nevertheless be noted that, even if we were only interested in theorems over $\mathbf{Q}$, we are forced to prove a modularity lifting theorem for any totally real field F (and prime p which splits completely in it). This is because we need to employ Taylor's Ihara avoidance technique [Tay08] to deal with issues of level raising and lowering at places away from p, and this step crucially relies on using solvable base change. We can then combine this modularity lifting result with base change techniques and the Moret-Bailly argument to achieve residual potential modularity, in order to prove our main potential modularity theorem.

In the light of the above discussion, in order to prove a modularity lifting theorem for Hilbert–Siegel modular forms it is natural to consider (a suitable localization of) either the cohomology complex $\mathrm{R}\Gamma(X, \omega^2)$ where X is an integral model over $\mathbf{Z}_p$ of the Hilbert–Siegel space, or of the ordinary part of the cohomology complex for a subspace of X obtained from the p-rank stratification. The required inputs for the modified Taylor–Wiles method are now:

(1) to prove that the cohomology is only supported in degrees 0 to $[F : \mathbf{Q}]$, and
(2) to prove local-global compatibility for the cohomology classes.

It is to some extent possible to solve (1) using higher Hida theory (although there are some issues), but (2) seems to be a more serious problem because we only know how to prove that the Galois representations associated to torsion classes in H^i satisfy the right local-global compatibility condition at $v|p$ if $i = 0$. Accordingly, we are unable to argue directly with such complexes.

Let the number of non-zero degrees of cohomology of the spaces we are considering be l_0+1; we refer to l_0 as the *defect*. (The original Taylor–Wiles method only applies if $l_0 = 0$, while if $l_0 > 0$ we use the method of [CG18]. As mentioned above, l_0 also has a Galois-theoretic interpretation: the sum of the dimensions of the local deformation rings is l_0 less than the corresponding dimension in the defect 0 case.) One key trick we employ in this paper is to reduce to situations where we only have to consider cohomology in at most two degrees (so the defect is at most one), i.e. it suffices to work with complexes consisting of at most two terms. This is where we take advantage of the product situation at p (because p splits in the totally real field). (Implicitly, what happens in this case is that any cohomology occurring in H^1 can also be seen via the Bockstein homomorphism as coming from H^0, *provided* that the characteristic zero classes in H^1 are also seen by the characteristic zero classes in H^0, and this can be established by automorphic considerations; so we only have to prove local-global compatibility for H^0.) We now explain how we do this in slightly more detail.

We assume that A has ordinary good reduction at all places $v|p$, so that

$$\rho_{A,p}|_{G_{F_v}} \simeq \begin{pmatrix} \lambda_{\alpha_v} & 0 & * & * \\ 0 & \lambda_{\beta_v} & * & * \\ 0 & 0 & \lambda_{\beta_v}^{-1}\varepsilon^{-1} & 0 \\ 0 & 0 & 0 & \lambda_{\alpha_v}^{-1}\varepsilon^{-1} \end{pmatrix},$$

where we furthermore assume that $\overline{\alpha}_v \neq \overline{\beta}_v$.

Although we expect that there should be a weight 2 eigenform associated to A of spherical level at p (because A has good reduction at p), it turns out that because A is ordinary at p, it is more natural to look for an eigenform f associated to A of Klingen level at p. The Klingen level structure is given by choosing a subgroup of order p inside $A[v]$ for all $v|p$. At Klingen level at v, there is a Hecke operator $U_{\mathrm{Kli}(v),1}$ whose eigenvalue on f should be $\alpha_v + \beta_v$, and a second Hecke operator $U_{\mathrm{Kli}(v),2}$ whose eigenvalue should be $\alpha_v\beta_v$. We observe that the second operator has an invertible eigenvalue (we say that f is Klingen ordinary) and this corresponds to the fact that the Galois representation $\rho_{A,p}|_{G_{F_v}}$ is ordinary.

There is another level structure that plays a role: the Iwahori level structure given by choosing a complete self dual flag of subgroups inside $A[v]$. For each $v|p$, there are two degeneracy maps from Iwahori level to Klingen level, and there are Hecke operators $U_{\mathrm{Iw}(v),1}, U_{\mathrm{Iw}(v),2} = U_{\mathrm{Kli}(v),2}$ at Iwahori level. Pulling back the expected form f by the degeneracy maps should yield eigenforms at Iwahori level which have eigenvalues α_v and β_v for $U_{\mathrm{Iw}(v),1}$ (we call them Iwahori ordinary).

We now return to the question of using modularity lifting theorems to find f. First of all, modularity lifting theorems with p-adic ordinary modular forms (i.e. with $l_0 = 0$) allow us to construct $2^{[F:\mathbf{Q}]}$ Iwahori ordinary p-adic modular forms whose eigenvalue for $U_{\mathrm{Iw}(v),1}$ is α_v or β_v, and whose eigenvalue for $U_{\mathrm{Kli}(v),2}$ is $\alpha_v\beta_v$. We suspect that these forms are classical, but as explained before, we don't know how to establish any geometric relation between them.

As a second step we apply a modularity lifting theorem in the case that the defect l_0 equals one. Let us isolate a place $v|p$. Using higher Hida theory, we construct a perfect complex of amplitude $[0,1]$ which is obtained by taking the ordinary (more precisely Iwahori ordinary at $w \neq v$, Klingen ordinary at v) cohomology of the open subspace of the Hilbert–Siegel Shimura variety which is ordinary and carries

an Iwahori level structure at all places $w \neq v$, and has p-rank at least one at v and carries a Klingen level structure.

We manage to prove that this cohomology carries a Galois representation which has the following type of local-global compatibility property:

(1) For all places $w|p$, $w \neq v$:

$$\rho_{A,p}|_{I_{F_w}} \simeq \begin{pmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & \varepsilon^{-1} & * \\ 0 & 0 & 0 & \varepsilon^{-1} \end{pmatrix}.$$

(2) For v:

$$\rho_{A,p}|_{I_{F_v}} \simeq \begin{pmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & \varepsilon^{-1} & 0 \\ 0 & 0 & 0 & \varepsilon^{-1} \end{pmatrix}.$$

Using the methods of [CG18], we can prove a modularity lifting theorem, and produce $2^{[F:\mathbf{Q}]-1}$ p-adic modular forms (which converge a lot more in the v direction) whose eigenvalue for $U_{\mathrm{Iw}(w),1}$ is α_w or β_w if $w \neq v$, and whose eigenvalue for $U_{\mathrm{Kli}(v),1}$ is $\alpha_v + \beta_v$, and whose eigenvalue for $U_{\mathrm{Kli}(w),2} = U_{\mathrm{Iw}(w),2}$ is $\alpha_w\beta_w$ for all $w|p$.

Our last step is to prove lots of linear relations between all these forms we have constructed. This step ultimately relies upon an abstract multiplicity one result which we prove using the Taylor–Wiles method. Exploiting these linear relations and using étale descent techniques, we first manage to construct a Klingen ordinary weight 2 modular form defined on the open subspace of the Hilbert–Siegel Shimura variety which has p-rank at least one at all $v|p$ and carries a Klingen level structure. We then manage, using analytic continuation techniques, to prove that this form extends to the full Shimura variety and is therefore classical.

1.3. **An outline of the paper.** We briefly explain the outline of the paper; we refer the reader to the introductions to the individual sections for a further explanation of their contents, and for some elaborations on the overview of our arguments above.

In §2 we recall some more or less standard background material on Galois representations, the local Langlands correspondence, local representation theory, and related topics. §3 discusses the Shimura varieties which we use, and some properties of their integral models and compactifications, and recalls the approach to the normalization of Hecke operators on coherent cohomology via cohomological correspondences which was introduced in [Pil20].

In §4 we construct the Hida complexes that we work with, and prove some of their basic properties (in particular, we prove that they are perfect complexes). In §5 we establish the "doubling" results that we will later use to prove local–global compatibility for Hilbert–Siegel modular forms over torsion rings. The basic strategy (employed in a number of other places, see [Gro90, Edi92, Wie14, CG18, CG20]) is to show that we can embed (via degeneracy maps) two copies of our space of ordinary modular forms at Klingen level into a space of ordinary modular forms of Iwahori level. This allows us to show that the corresponding Galois representations are ordinary (in the Iwahori sense) in two different ways, namely, with α_w and β_w as unramified subspaces. Then the genericity assumption $\overline{\alpha}_w \neq \overline{\beta}_w$ forces there to

be a 2-dimensional unramified summand of our representation. The key technical difficulty is proving that the direct sum of the degeneracy maps does indeed give an embedding. All previous incarnations of the doubling phenomenon ultimately relied on the q-expansion principle, but our argument is more geometric, and ultimately rests on analyzing the effect of the Hecke operator $Z_w = U_{\mathrm{Kli}(w),1} - U_{\mathrm{Iw}(w),1}$ along the w-non-ordinary locus.

In §6 we prove that a characteristic zero classicality result for the H^0 of our Hida complexes, using Coleman theory. We also show that the complexes we consider are balanced, in the sense that they have Euler characteristic zero, using a somewhat intricate interplay between three objects — the complex of classical forms, the complex of overconvergent forms, and our complex of (Klingen) ordinary forms.

In §7 we carry out our main Taylor–Wiles patching arguments in the cases that $l_0 = 0$ and $l_0 = 1$. We then prove our main modularity lifting theorem in §8, using analytic continuation, étale descent, and linear algebra arguments based on the doubling results of §5 to reduce to the classicality results of §6.

In §9, we apply our main automorphy lifting theorem to prove the potential automorphy of abelian surfaces. The basic idea is to use a version of the p-q trick (first employed by Wiles as the 3-5 trick), together with an application of a theorem of Moret-Bailly, to connect general abelian surfaces via a chain of congruences to the restriction of scalars of an elliptic curve over a totally real quadratic extension of F, which we know already by [Tay02] to be potentially modular. We are also left to deal directly with some cases of abelian surfaces with small Mumford–Tate groups, which can mostly be done immediately with an appeal to the theory of Grossencharacters. We also include a number of applications as mentioned in the introduction, including elliptic curves over quadratic extensions of F.

In §10, we give applications to the automorphy of abelian surfaces. We show that, given any mod 3 representation $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_3)$ with (inverse) cyclotomic similitude character, it can be realized (in infinitely many ways) as the 3-torsion of an abelian surface over $\mathbf{Q}$. Here we exploit some classical geometry related to the Burkhardt quartic, which is isomorphic to a compactification of $\mathcal{A}_2(3)$. The key point is to show that the variety given by the twist of $\mathcal{A}_2(3)$ by $\overline{\rho}$ has sufficiently many rational points. We do this by proving it is unirational over $\mathbf{Q}$ via a map of degree at most 6. The argument is similar to that of [SBT97], except that it is applied not to the twist of $\mathcal{A}_2(3)$ itself but to a twist of a degree 6 rational cover, which has the pleasing property (unlike the Burkhardt quartic itself) that the birational map to $\mathbf{P}^3$ over $\mathbf{Q}$ can be made equivariant with respect to the action of the automorphism group $\mathrm{PSp}_4(\mathbf{F}_3)$. Finally, we conclude with a discussion of the paramodular conjecture and its relationship to the standard conjectures, and explain why the original formulation of this conjecture requires a minor modification.

1.4. **Some further remarks.** For length reasons, we did not try to optimize all of our theorems — for example, our arguments would surely extend to prove the potential automorphy of some GSp_4-type abelian varieties, but sticking with abelian surfaces makes the Moret-Bailly arguments somewhat simpler, and (by using a trick) we manage to avoid any character building whatsoever. However, we have gone to some lengths to treat the case $p = 3$, and to use a weaker notion of p-distinguishedness than in [CG20]; while this is not necessary for our applications to potential modularity, it significantly increases the applicability of our theorems to actual modularity problems.

1.4.1. *The work of Arthur.* It should be noted that we use Arthur's multiplicity formula for the discrete spectrum of GSp_4, as announced in [Art04]. A proof of this (relying on Arthur's work for symplectic and orthogonal groups in [Art13]) was given in [GT19], but this proof is only as unconditional as the results of [Art13] and [MW16a, MW16b]. In particular, it depends on cases of the twisted weighted fundamental lemma that were announced in [CL10], but whose proofs have not yet appeared, as well as on the references [A24], [A25], [A26] and [A27] in [Art13], which at the time of writing have not appeared publicly.

1.4.2. *Curves of higher genus.* One may well ask whether the methods of this paper could be used to prove (potential) modularity of curves of genus $g \geq 3$ whose Jacobians have trivial endomorphism rings. At the moment, this seems exceedingly unlikely without some substantial new idea. All generalizations of the Taylor–Wiles method to this point require that the automorphic representations in question are associated to the Betti cohomology groups of locally symmetric spaces, or the coherent cohomology groups of Shimura varieties, which have integral structures and hence allow one to talk about *congruences* between automorphic forms. Symplectic motives of rank $2g$ over $\mathbf{Q}$ are conjecturally associated to automorphic representations for the (split) orthogonal group SO_{2g+1} (when $g = 1$ or $g = 2$, there are well-known exceptional isomorphisms which allow us to replace SO_{2g+1} by the groups GL_2 and GSp_4 respectively). Following [BK14], Gross has made some precise conjectures concerning the level structures of newforms associated to such conjectural automorphic representations in [Gro16].

The automorphic representations contributing to the Betti cohomology groups of locally symmetric spaces have regular infinitesimal characters, so can only be used for $g = 1$. The automorphic representations contributing to the coherent cohomology of orthogonal Shimura varieties are representations of the inner form $\mathrm{SO}(2g-1, 2)$ of SO_{2g+1} (which is non-split if $g > 1$), whose infinity components π_∞ are furthermore either discrete series, or non-degenerate limits of discrete series.

If $g = 1$, the representations considered by Gross in [Gro16] are discrete series, and if $g = 2$, they are non-degenerate limits of discrete series, but if $g \geq 3$, then neither possibility occurs, so the automorphic representations do not contribute to the cohomology (of any kind) of the corresponding Shimura variety. (Another way of seeing this is to compute the possible infinitesimal characters of the automorphic representations corresponding to automorphic vector bundles on the Shimura variety, or equivalently the Hodge–Tate weights of the expected $2g$-dimensional symplectic Galois representations; one finds that no Hodge–Tate weight can occur with multiplicity bigger than 2, while the symplectic Galois representations coming from the étale H^1 of a curve of genus g have weights $0, 1$ each occurring with multiplicity g.) In particular, the general modularity problem for curves of genus $g \geq 3$ seems at least as hard as proving non-solvable cases of the Artin conjecture for totally even representations, and even proving the modularity of a single such curve with Mumford–Tate group GSp_{2g} seems completely out of reach.

On the other hand, there are some special families in higher genus which may well be amenable to our method. In particular, the Tate module of a cyclic trigonal genus three curve (so-called Picard curves, with affine equations of the form $y^3 = x^4 + ax^2 + bx + c$) defined over $\mathbf{Q}$ splits (over $\mathbf{Q}(\sqrt{-3})$) into two essentially conjugate self-dual irregular 3-dimensional representations of $G_{\mathbf{Q}(\sqrt{-3})}$. These

Galois representations conjecturally correspond (see the appendix to [Til06a]) to automorphic representations π for a form of $U(2,1)/\mathbf{Q}$ (splitting over $\mathbf{Q}(\sqrt{-3})$) such that π_∞ is a non-degenerate limit of discrete series and contributes to the coherent cohomology of the associated Shimura variety. The methods of this paper should apply (in principle) to these curves.

1.4.3. K3 *surfaces.* Our results should also have applications to the Hasse–Weil conjecture for K3 surfaces over totally real fields with geometric Picard number ≥ 17. While we do not undertake a detailed study of this problem here, we discuss it in §9.4.

1.4.4. *A comparison of this paper with* [ACC+18]. It follows from Theorem 1.1.3 that any elliptic curve E over a CM field K/F is potentially modular (simply consider the abelian surface given by Weil restriction of scalars of E from K to F). This result is also proved in [ACC+18]. Perhaps surprisingly, there is relatively little overlap between the two proofs. For example, our argument does not require any of the results of Scholze [Sch15] on the construction of Galois representations, nor the derived version of Ihara avoidance required in [ACC+18]. The only common theme is the use of the modified Taylor–Wiles method of [CG18]. To further illustrate the difference, it is also proved in [ACC+18] that the nth symmetric power of any such E is potentially automorphic, which is not directly accessible from our approach. On the other hand, we also deduce (Theorem 1.1.4) the potential modularity of elliptic curves over fields like $F = \mathbf{Q}(\sqrt[4]{2})$, which seems out of reach using the methods of [ACC+18].

1.5. **Acknowledgements.** We would like to thank all of the referees for their numerous helpful comments and corrections. Both David Geraghty and Jacques Tilouine have made important contributions to the problem of modularity for abelian surfaces; we would especially like to thank them for many helpful discussions over the years. We would also like to thank Patrick Allen, Kevin Buzzard, Brian Conrad, Matthew Emerton, Najmuddin Fakhruddin, Dick Gross, Robert Guralnick, Florian Herzig, Christian Johansson, Keerthi Madapusi Pera, Rutger Noot, Madhav Nori, Ralf Schmidt, Olivier Taïbi, Richard Taylor, Jack Thorne, Andrew Wiles, and Liang Xiao for helpful conversations. We finally want to thank Bruno Klingler and the University of Paris 7 for hosting us during part of this project.

2. Background material

In this section we recall a variety of more or less well-known results that we will use in the body of the paper.

2.1. Notation and conventions.

2.1.1. GSp_4. We define GSp_4 to be the reductive group over $\mathbf{Z}$ defined as a subgroup of GL_4 by

$$\mathrm{GSp}_4(R) = \{g \in \mathrm{GL}_4(R) : gJg^t = \nu(g)J\}$$

where $\nu(g)$ is the similitude factor (which is uniquely determined by g, and which we sometimes call the multiplier factor), and J is the antisymmetric matrix

$$\begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix}$$

where $s = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Note that the map $\nu : g \mapsto \nu(g)$ is a homomorphism $\mathrm{GSp}_4 \to \mathbf{G}_m$.

We let Sp_4 be the subgroup with $\nu = 1$, and we let $B \subset G = \mathrm{GSp}_4$ be the Borel subgroup of upper triangular matrices, and $T \subset B$ be the diagonal maximal torus. Write $W_G = N_G(T)/T$ for the Weyl group of (G, T). It acts on the character group via $w \cdot \lambda(t) = \lambda(w^{-1}tw)$. It is generated by $s_1 = \begin{pmatrix} s & 0_2 \\ 0_2 & s \end{pmatrix}$ and $s_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & s' & 0 \\ 0 & 0 & 1 \end{pmatrix}$ where $s' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and admits the presentation

$$W_G = \langle s_1, s_2 | s_1^2 = s_2^2 = (s_1 s_2)^4 = 1 \rangle.$$

Write $X^*(T)$ (resp. $X_*(T)$) for the group of characters (resp. cocharacters) of T. We identify $X^*(T)$ with the lattice in $\mathbf{Z}^3$ of triples $(a, b; c) \in \mathbf{Z}^3$ such that $c \equiv a + b \pmod 2$ via

$$\lambda : t = \mathrm{diag}(t_1, t_2, \nu t_2^{-1}, \nu t_1^{-1}) \mapsto t_1^a t_2^b \nu^{(c-a-b)/2}.$$

In particular, the central character is given by $\lambda(\mathrm{diag}(z, z, z, z)) = z^c$. The simple roots are $\alpha_1 = (1, -1; 0)$ and $\alpha_2 = (0, 2; 0)$; α_1 is the short root. Note that the α_i determine the reflections s_i. The similitude factor is $(0, 0; 2)$.

The root datum (G, B, T) determines the dual root datum $(\widehat{G}, \widehat{B}, \widehat{T})$, where $\widehat{G}$ is the dual group GSpin_5. We always identify GSpin_5 with GSp_4 via the spin isomorphism (see for example [MT02, §3.2] for a detailed explanation of this). In particular, the cocharacter in $X_*(\widehat{T})$ corresponding to the character $(a, b; c) \in X^*(T)$ defined above is given by

$$t \mapsto \mathrm{diag}(t^{(a+b+c)/2}, t^{(a-b+c)/2}, t^{(-a+b+c)/2}, t^{(-a-b+c)/2}).$$

We write $\mathfrak{g}$ and $\mathfrak{b}$ for the Lie algebras of GSp_4 and B, and $\mathfrak{g}^0$ and $\mathfrak{b}^0$ for the Lie algebras of Sp_4 and $B \cap \mathrm{Sp}_4$. If v is a finite place of a number field F, with residue field $k(v)$, then we have the standard parahoric subgroups of $\mathrm{GSp}_4(F_v)$:

- The hyperspecial subgroup $\mathrm{GSp}_4(\mathcal{O}_{F_v})$.
- The paramodular subgroup $\mathrm{Par}(v)$, the stabilizer in $\mathrm{GSp}_4(F_v)$ of $\mathcal{O}_{F_v} \oplus \mathcal{O}_{F_v} \oplus \mathcal{O}_{F_v} \oplus \varpi_v \mathcal{O}_{F_v}$, where $\varpi_v \in \mathcal{O}_{F_v}$ is a uniformizer.
- The Siegel parahoric $\mathrm{Si}(v)$, the preimage in $\mathrm{GSp}_4(\mathcal{O}_{F_v})$ of those matrices in $\mathrm{GSp}_4(k(v))$ of the form
$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}.$$
- the Klingen parahoric $\mathrm{Kli}(v)$, the preimage in $\mathrm{GSp}_4(\mathcal{O}_{F_v})$ of those matrices in $\mathrm{GSp}_4(k(v))$ of the form
$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$
- the Iwahori subgroup $\mathrm{Iw}(v)$, the preimage of $B(k(v))$ in $\mathrm{GSp}_4(\mathcal{O}_{F_v})$.

2.1.2. *Algebra.* If R is a local ring we write $\mathfrak{m}_R$ for the maximal ideal of R.

If M is a perfect field, we let $\overline{M}$ denote an algebraic closure of M and G_M the absolute Galois group $\mathrm{Gal}(\overline{M}/M)$. For each prime p not equal to the characteristic of M, we let ε_p denote the p-adic cyclotomic character and $\overline{\varepsilon}_p$ its reduction modulo p. We will usually drop p from the notation and simply write ε, $\overline{\varepsilon}$.

If K is a finite extension of $\mathbf{Q}_p$ for some p, we write K^{nr} for its maximal unramified extension; I_K for the inertia subgroup of G_K; $\mathrm{Frob}_K \in G_K/I_K$ for the geometric Frobenius; and W_K for the Weil group. If L/K is a Galois extension we will write $I_{L/K}$ for the inertia subgroup of $\mathrm{Gal}(L/K)$. We will write $\mathrm{Art}_K : K^\times \xrightarrow{\sim} W_K^{\mathrm{ab}}$ for the Artin map normalized to send uniformizers to geometric Frobenius elements.

If ρ is a continuous representation of G_K over $\overline{\mathbf{Q}}_l$ for some $l \neq p$, valued either in some GL_n or in GSp_4, then we write $\mathrm{WD}(\rho)$ for the corresponding Weil–Deligne representation. (By definition, a GSp_4-valued Weil–Deligne representation is just a GSp_4-valued representation of the Weil–Deligne group, i.e. it is considered up to GSp_4-conjugacy). If ρ is a de Rham representation of G_K on a $\overline{\mathbf{Q}}_p$-vector space W, then we will write $\mathrm{WD}(\rho)$ for the corresponding Weil–Deligne representation of W_K, and if $\tau : K \hookrightarrow \overline{\mathbf{Q}}_p$ is a continuous embedding of fields, then we will write $\mathrm{HT}_\tau(\rho)$ for the multiset of Hodge–Tate numbers of ρ with respect to τ, which by definition contains i with multiplicity $\dim_{\overline{\mathbf{Q}}_p}(W \otimes_{\tau,K} \widehat{\overline{K}}(i))^{G_K}$. Thus, for example, $\mathrm{HT}_\tau(\varepsilon) = \{-1\}$.

Let $K/\mathbf{Q}$ be a finite extension. If v is a finite place of K we write $k(v)$ for its residue field, q_v for $\#k(v)$, and Frob_v for Frob_{K_v}. If v is a real place of K, then we will let $[c_v]$ denote the conjugacy class in G_K consisting of complex conjugations associated to v.

We will frequently adopt the following notation: we let $p > 2$ be prime, and we let E be a finite extension of $\mathbf{Q}_p$ with ring of integers $\mathcal{O}$, uniformizer λ and residue field k.

We will sometimes use the following well-known lemma without comment.

Lemma 2.1.3. *Let Γ be a group and let L be an algebraically closed field. Then a semisimple representation $\Gamma \to \mathrm{GSp}_4(L)$ is determined up to conjugacy by the composite $\Gamma \to \mathrm{GSp}_4(L) \to \mathrm{GL}_4(L) \times \mathrm{GL}_1(L)$, where the second factor records the similitude character.*

Proof. This follows (for example) from the proof of Lemma 6.1 of [GT11a]. □

2.1.4. *Galois cohomology.* If L/K is an extension of fields, k is a field, and V is a finite-dimensional k-vector space with an action of $\mathrm{Gal}(L/K)$, then we write $H^i(L/K, V)$ for $H^i(\mathrm{Gal}(L/K), V)$, and $h^i(L/K, V)$ for $\dim_k H^i(L/K, V)$. We write $H^i(K, V)$ and $h^i(K, V)$ for $H^i(\overline{K}/K, V)$ and $h^i(\overline{K}/K, V)$ respectively.

2.1.5. *Automorphic representations.* We will use the letter π for automorphic representations of GSp_4, Π for automorphic representations of GL_n (usually with $n = 4$), and $\boldsymbol{\pi}$ for automorphic representations of GL_2. We decorate these in various ways, and aim to be consistent in such decorations. For example, Π will usually denote the transfer to GL_4 of π in the sense of §2.9, so that for example Π_2' will denote the transfer of π_2'.

2.2. **Induction of two-dimensional representations.** We will sometimes want to induce representations from GL_2 to GSp_4. Suppose that K/F is a quadratic

extension of fields, and that $r : G_K \to \mathrm{GL}_2(L)$ is a representation, for some field L. Choose $\sigma \in G_F \setminus G_K$, and assume that $\det r$ extends to a character χ of G_F Let $\rho := \mathrm{Ind}_{G_K}^{G_F} r : G_F \to \mathrm{GL}_4(L)$. The representation $\wedge^2\rho$ admits the characters χ and $\chi \otimes \eta_{F/K}$ as constituents, where $\eta_{F/K}$ denotes the quadratic character. In particular, the representation ρ generally preserves two symplectic forms, and hence gives rise to two representations $\rho_1, \rho_2 : G_F \to \mathrm{GSp}_4(L)$ with similitude factors χ and $\chi \otimes \eta_{F/K}$ respectively. To describe these more explicitly, let V denote a model for r so that $W = V \oplus \sigma V$ is a model for ρ. Then the Galois action of W preserves (up to scalar) the symplectic form given by choosing an arbitrary non-degenerate symplectic form on V, letting σV and V be orthogonal, and then defining $\sigma v_1 \wedge \sigma v_2$ consistently to be either $\chi(\sigma) v_1 \wedge v_2$ or $-\chi(\sigma) v_1 \wedge v_2 = \chi \otimes \eta_{F/K}(\sigma) v_1 \wedge v_2$. The image of $(A, B) \in \mathrm{GL}_2(E) \times \mathrm{GL}_2(E)$ with $\det(A) = \det(B)$ inside GSp_4 relative to our choice of J can be given by

$$\begin{pmatrix} * & 0 & 0 & * \\ 0 & * & * & 0 \\ 0 & * & * & 0 \\ * & 0 & 0 & * \end{pmatrix} \cap \mathrm{GSp}_4(E).$$

In our applications, it will always be the case that $\det r$ is the inverse of the cyclotomic character of G_K, and we will write simply write $\mathrm{Ind}_{G_K}^{G_F} r$ for the corresponding symplectic representation with similitude factor the inverse of the cyclotomic character of G_F. For example, if K/F is a quadratic extension of number fields, E is an elliptic curve over K, and r is the dual of the p-adic Tate module of E, then $\mathrm{Ind}_{G_K}^{G_F} r$ is the dual of the p-adic Tate module of the abelian surface $A = \mathrm{Res}_{K/F} E$, and the corresponding symplectic structure on this representation coincides with the one coming from the Weil pairing on A. This is because the representation on the Tate module of A is the induction of the corresponding representation on the Tate module of E, and because the similitude character on the Tate module of an abelian variety is always given by the cyclotomic character.

2.3. **The non-archimedean local Langlands correspondence.** Let $K/\mathbf{Q}_l$ be a finite extension for some l. We will let rec_K be the local Langlands correspondence of [HT01], so that if π is an irreducible complex admissible representation of $\mathrm{GL}_n(K)$, then $\mathrm{rec}_K(\pi)$ is a Frobenius semi-simple Weil–Deligne representation of the Weil group W_K. We will write rec for rec_K when the choice of K is clear.

If (r, N) is a Weil–Deligne representation of W_K we will write $(r, N)^{F-\mathrm{ss}}$ for its Frobenius semisimplification. If π_i is an irreducible smooth representation of $\mathrm{GL}_{n_i}(K)$ for $i = 1, 2$ we will write $\pi_1 \boxplus \pi_2$ for the irreducible smooth representation of $\mathrm{GL}_{n_1+n_2}(K)$ with $\mathrm{rec}(\pi_1 \boxplus \pi_2) = \mathrm{rec}(\pi_1) \oplus \mathrm{rec}(\pi_2)$. If L/K is a finite extension and if π is an irreducible smooth representation of $\mathrm{GL}_n(K)$ we will write $\mathrm{BC}_{L/K}(\pi)$ for the base change of π to L which is characterized by $\mathrm{rec}_L(\mathrm{BC}_{L/K}(\pi)) = \mathrm{rec}_K(\pi)|_{W_L}$.

We denote the local Langlands correspondence of [GT11a] by $\mathrm{rec}_{\mathrm{GT}}$; this is a surjective finite-to-one map from the set of equivalence classes of irreducible smooth complex representations of $\mathrm{GSp}_4(K)$ to the set of GSp_4-conjugacy classes of $\mathrm{GSp}_4(\mathbf{C})$-valued Weil–Deligne representations of W_K, which we normalize so that $\mathrm{rec}_{\mathrm{GT}}(\pi \otimes (\chi \circ \nu)) = \mathrm{rec}_{\mathrm{GT}}(\pi) \otimes \mathrm{rec}(\chi)$, and $\nu \circ \mathrm{rec}_{\mathrm{GT}}(\pi) = \mathrm{rec}(\omega_\pi)$, where ω_π is the central character of π.

We fix once and for all for each prime p an isomorphism $\imath = \imath_p : \mathbf{C} \cong \overline{\mathbf{Q}}_p$. We will generally omit these isomorphisms from our notation, in order to avoid clutter.

In particular, we will frequently use that $\imath$ determines a square root of p in $\overline{\mathbf{Q}}_p$ (corresponding to the positive square root of p in $\mathbf{C}$). We write rec_p and $\mathrm{rec}_{\mathrm{GT},p}$ for the local Langlands correspondences for $\overline{\mathbf{Q}}_p$-representations given by conjugating by $\imath$. These depend on $\imath$, but in practice this does not cause us any difficulty; see Remark 2.3.2.

Definition 2.3.1. If $\rho : G_K \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ is a continuous representation for some $p \neq l$, then we write $L(\rho)$ for the L-packet associated to ρ, which by definition is the set of equivalence classes of irreducible smooth $\overline{\mathbf{Q}}_p$-representations π of $\mathrm{GSp}_4(K)$ with the property that $\mathrm{rec}_{\mathrm{GT},p}(\pi \otimes |\nu|^{-3/2}) \cong \mathrm{WD}(\rho)^{F-\mathrm{ss}}$.

(In accordance with the convention explained above, note that $|\nu|^{-3/2}$ makes sense because we have a fixed square root of p.)

Remark 2.3.2. It is presumably possibly to show that the twist of $\mathrm{rec}_{\mathrm{GT}}$ in Definition 2.3.1 (which will be present whenever we consider $\mathrm{rec}_{\mathrm{GT},p}$) gives a local Langlands correspondence for $\overline{\mathbf{Q}}_p$-representations which is independent of the choice of $\imath$, but we have not tried to establish this, as we do not need it. We make (implicit) use of this for unramified representations, and of the statement that the rank of the monodromy operator associated to a representation with Iwahori-fixed vectors is independent of the choice of $\imath$, both of which are easily verified explicitly.

Remark 2.3.3. We will from now on usually regard automorphic representations as being defined over $\overline{\mathbf{Q}}_p$, rather than $\mathbf{C}$, by means of the fixed isomorphism $\imath : \mathbf{C} \cong \overline{\mathbf{Q}}_p$. We will not in general draw attention to this, and no confusion should arise on the few occasions (for example, when considering compatible systems) where we think of them as being over $\mathbf{C}$.

If L/K is a finite solvable Galois extension of number fields and if π is a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbf{A}_K)$, we will write $\mathrm{BC}_{L/K}(\pi)$ for its base change to L (which exists by the main results of [AC89]), an (isobaric) automorphic representation of $\mathrm{GL}_n(\mathbf{A}_L)$ satisfying

$$\mathrm{BC}_{L/K}(\pi)_w = \mathrm{BC}_{L_w/K_v}(\pi_v)$$

for all places w of L where $v = w|_K$ is the restriction of w to K. If π_i is an automorphic representation of $\mathrm{GL}_{n_i}(\mathbf{A}_K)$ for $i = 1, 2$ we will write $\pi_1 \boxplus \pi_2$ for the automorphic representation of $\mathrm{GL}_{n_1+n_2}(\mathbf{A}_K)$ satisfying

$$(\pi_1 \boxplus \pi_2)_v = \pi_{1,v} \boxplus \pi_{2,v}$$

for all places v of K.

If (r, N) is a Weil–Deligne representation, then we write $n((r, N))$ for the rank of N. If π is an irreducible admissible representation of $\mathrm{GL}_n(K)$ (resp. $\mathrm{GSp}_4(K)$), then we write $n(\pi)$ for $n(\mathrm{rec}(\pi))$ (resp. $n(\mathrm{rec}_{\mathrm{GT}}(\pi))$).

2.4. **Local representation theory.** In this section, we recall a number of more or less well-known results about the representation theory of $\mathrm{GSp}_4(K)$, where K is a local field of characteristic zero. Some of these results are in [GT05], but for convenience we have gathered them all together here, and have usually given proofs. Since our applications of this material are all global, and some of the definitions we make (such as the normalizations of Hecke operators at places dividing p) depend on global information, we have chosen to work in the same global setting that we consider in the rest of the paper.

Let $p > 2$ be prime, and let F be a totally real field in which p splits completely. Let $E/\mathbf{Q}_p$ be a finite extension with ring of integers $\mathcal{O}$ and residue field k. Let v be a finite place of F, and fix a uniformizer $\varpi_v \in \mathcal{O}_{F_v}$. For most of this section, we will allow v to divide p, although at the end of the section, we will prove some results (which follow those of [KT17] for GL_n) under the assumption that $q_v \equiv 1 \pmod p$. We fix once and for all a square root $q_v^{1/2} \in E$.

2.4.1. *Generalities.* We begin by recalling some results on Iwahori Hecke algebras. It costs us nothing to recall these in a more general setting, so we temporarily let $G/\mathcal{O}_{F_v}$ be a split reductive group with $T \subset B = T \cdot U$ a maximal torus and Borel (with unipotent radical U), and let N be the normalizer of T in G. Let $W = N(F_v)/T(F_v)$ be the Weyl group. Let $\Delta \subset X^*(T)$ be the simple roots. We write $\widetilde{W} = N(F_v)/T(\mathcal{O}_{F_v})$ for the extended affine Weyl group.

Let $\mathrm{Iw}(v) = \ker(G(\mathcal{O}_{F_v}) \to B(k(v)))$ be an Iwahori subgroup, and let $\mathrm{Iw}_1(v) = \ker(G(\mathcal{O}_{F_v}) \to U(k(v)))$ be a pro-v Iwahori subgroup. Let

$$\mathcal{H}_1 = \mathcal{H}_1(v) = \mathcal{O}[\mathrm{Iw}_1(v)\backslash G(F_v)/\mathrm{Iw}_1(v)] = \mathcal{O}[G(F_v)//\mathrm{Iw}_1(v)]$$

be the pro-v Iwahori Hecke algebra. (Here $G//K$ denotes $K\backslash G/K$ — we tend to prefer the first notation but we also sometimes use the second notation since it is more compact and some of our expressions are already typographically somewhat complicated.)

We let $T(\mathcal{O}_{F_v})_1 = (\ker T(\mathcal{O}_{F_v}) \to T(k(v)))$. We also let

$$T(F_v)^+ = \{x \in T(F_v) \mid \alpha(x) \in \mathcal{O}_{F_v}, \forall \alpha \in \Delta\}.$$

For $g \in G(F_v)$, we write $[\mathrm{Iw}_1(v) g \mathrm{Iw}_1(v)] \in \mathcal{H}_1$ for the characteristic function of the double coset $\mathrm{Iw}_1(v) g \mathrm{Iw}_1(v)$.

Proposition 2.4.2. *For $x, y \in T(F_v)^+$, we have*

$$[\mathrm{Iw}_1(v) x \mathrm{Iw}_1(v)] \cdot [\mathrm{Iw}_1(v) y \mathrm{Iw}_1(v)] = [\mathrm{Iw}_1(v) xy \mathrm{Iw}_1(v)]$$

and moreover $[\mathrm{Iw}_1(v) x \mathrm{Iw}_1(v)] \in (\mathcal{H}_1[1/p])^\times$. If $v \nmid p$, then in fact $[\mathrm{Iw}_1(v) x \mathrm{Iw}_1(v)] \in \mathcal{H}_1^\times$.

Proof. The first statement is a special case of [Cas, Lem. 4.1.5], while the rest is immediate from [Vig05, Cor. 1]. □

As a result, there is a homomorphism

$$T(F_v) \to (\mathcal{H}_1[1/p])^\times$$

which is defined as follows: write $x \in T(F_v)$ as $x = yz^{-1}$ with $y, z \in T(F_v)^+$ and send x to

$$(\delta_B^{1/2}(y)[\mathrm{Iw}_1(v) y \mathrm{Iw}_1(v)])(\delta_B^{1/2}(z)[\mathrm{Iw}_1(v) z \mathrm{Iw}_1(v)])^{-1}$$

where δ_B is the modulus character. The kernel of this homomorphism is $T(\mathcal{O}_{F_v})_1$. If $v \nmid p$, then the image of the homomorphism is in $\mathcal{H}_1^\times$.

Proposition 2.4.3. *Let π be a smooth admissible $\overline{E}[G(F_v)]$-module. Then the map $\pi \to \pi_U$, where π_U is the (normalized) Jacquet module, induces an isomorphism of $\overline{E}[T(F_v)]$-modules*

$$\pi^{\mathrm{Iw}_1(v)} \to (\pi_U)^{T(\mathcal{O}_{F_v})_1}.$$

Proof. By [Cas, Lem. 4.1.1] (noting that the Jacquet module in this reference is *not* the normalized Jacquet module), the map $\pi \to \pi_U$ induces an $\overline{E}[T(F_v)]$-module homomorphism $\pi^{\mathrm{Iw}_1(v)} \to (\pi_U)^{T(\mathcal{O}_{F_v})_1}$. It is an isomorphism by [Cas, Prop. 4.1.4] and Proposition 2.4.2. □

For a character $\chi : T(F_v) \to \overline{E}^\times$, write $\pi(\chi) = \text{n-Ind}_{B(F_v)}^{G(F_v)}\chi$ for the corresponding principal series representation. Then we recall

Proposition 2.4.4. *For $\chi : T(F_v) \to \overline{E}^\times$ there is an isomorphism of $\overline{E}[T(F_v)]$-modules*

$$(\pi(\chi)_U)^{\mathrm{ss}} \simeq \bigoplus_{w \in W} \overline{E}(w \cdot \chi).$$

Proof. This is a special case of [Cas, Thm. 6.3.5]. □

We say that $\pi(\chi)$ is a *tame principal series* if χ is trivial on $T(\mathcal{O}_{F_v})_1$ and an *unramified principal series* if χ is trivial on $T(\mathcal{O}_{F_v})$. The results recalled above immediately imply the well-known facts that if π is an irreducible smooth $\overline{E}[G(F_v)]$-module, then $\pi^{\mathrm{Iw}_1(v)} \neq \{0\}$ if and only if π is a constituent of a tame principal series, and $\pi^{\mathrm{Iw}(v)} \neq \{0\}$ if and only if π is a constituent of an unramified principal series.

Write $\mathcal{H} := \mathcal{O}[\mathrm{Iw}(v)\backslash G(F_v)/\mathrm{Iw}(v)]$ for the Iwahori Hecke algebra. This enjoys similar properties to those of $\mathcal{H}_1$ recalled above; in particular, the analogue of Proposition 2.4.2 gives an embedding $E[X_*(T)] \hookrightarrow \mathcal{H}[1/p]$, and if $v \nmid p$, then this restricts to an embedding $\mathcal{O}[X_*(T)] \hookrightarrow \mathcal{H}$.

2.4.5. *Principal series for* GSp_4. We now specialize our discussion to $G = \mathrm{GSp}_4$. We recall some known results on constituents of unramified principal series representations; many of these results are originally due to [ST93], but for convenience we refer to the tables in [RS07b, App. A]. (Note that the compatibility of the proposed Langlands parameters in [RS07b, App. A.5] with the correspondence $\mathrm{rec}_{\mathrm{GT}}$ is proved in [GT11b, Prop. 13.1].)

If χ_1, χ_2, σ are characters of $F_v^\times$, then we write

$$\chi_1 \times \chi_2 \rtimes \sigma := \text{n-Ind}_{B(F_v)}^{\mathrm{GSp}_4(F_v)} \chi_1 \otimes \chi_2 \otimes \sigma,$$

where

$$\chi_1 \otimes \chi_2 \otimes \sigma : \begin{pmatrix} a & * & * & * \\ & b & * & * \\ & & cb^{-1} & * \\ & & & ca^{-1} \end{pmatrix} \mapsto \chi_1(a)\chi_2(b)\sigma(c).$$

Proposition 2.4.6.

(1) $\chi_1 \times \chi_2 \rtimes \sigma$ is irreducible if and only if none of $\chi_1, \chi_2, \chi_1\chi_2^{\pm 1}$ is equal to $|\cdot|_v^{\pm 1}$.
(2) If π is an irreducible constituent of $\chi_1 \times \chi_2 \rtimes \sigma$, then

$$\mathrm{rec}_{\mathrm{GT},p}(\pi)^{\mathrm{ss}} = \sigma \circ \mathrm{Art}_{F_v}^{-1} \otimes \left((\chi_1\chi_2) \circ \mathrm{Art}_{F_v}^{-1} \oplus \chi_1 \circ \mathrm{Art}_{F_v}^{-1} \oplus \chi_2 \circ \mathrm{Art}_{F_v}^{-1} \oplus 1\right).$$

(3) If $\chi_1 \times \chi_2 \rtimes \sigma$ is irreducible, then $\mathrm{rec}_{\mathrm{GT},p}(\chi_1 \times \chi_2 \rtimes \sigma)$ is semisimple (that is, $N = 0$).

Proof. Part (1) is [ST93, Lem. 3.2]. Parts (2) and (3) follow immediately from rows I–VI of [RS07b, Table A.7]. □

2.4.7. *Spherical Hecke operators.* Define matrices

$$\beta_{v,0} = \mathrm{diag}(\varpi_v, \varpi_v, \varpi_v, \varpi_v),$$
$$\beta_{v,1} = \mathrm{diag}(\varpi_v, \varpi_v, 1, 1),$$
$$\beta_{v,2} = \mathrm{diag}(\varpi_v^2, \varpi_v, \varpi_v, 1).$$

We have the spherical Hecke operators $T_{v,i} = [\mathrm{GSp}_4(\mathcal{O}_{F_v})\beta_{v,i}\,\mathrm{GSp}_4(\mathcal{O}_{F_v})]$, which are independent of ϖ_v. It is easy to check (using Proposition 2.4.6 (2)) that if π is an unramified representation of $\mathrm{GSp}_4(F_v)$ (that is, if $\pi^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \neq 0$, so that π is a constituent of an unramified principal series), then the characteristic polynomial of $\mathrm{rec}_{\mathrm{GT},p}(\pi \otimes |\nu|^{-3/2})(\mathrm{Frob}_v)$ is

$$Q_v(X) := X^4 - t_{v,1}X^3 + (q_v t_{v,2} + (q_v^3 + q_v)t_{v,0})X^2 - q_v^3 t_{v,0}t_{v,1}X + q_v^6 t_{v,0}^2, \tag{2.4.8}$$

where we are writing $t_{v,i}$ for the eigenvalue of the operator $T_{v,i}$ on $\pi^{\mathrm{GSp}_4(\mathcal{O}_{F_v})}$.

Definition 2.4.9. We say that the *Hecke parameters* of π are the roots of $Q_v(X)$, ordered in such a way that the pairs of roots $(1,4)$ and $(2,3)$ both multiply to give the value γ_v of the similitude character evaluated on Frob_v. We write these Hecke parameters as $[\alpha_v, \beta_v, \gamma_v\beta_v^{-1}, \gamma_v\alpha_v^{-1}]$, where implicitly we view these terms as labelling the vertices of a square:

$$\begin{tikzcd}
\alpha_v \arrow[r, dash] \arrow[d, dash] & \beta_v \arrow[d, dash] \\
\gamma_v\beta_v^{-1} \arrow[r, dash] & \gamma_v\alpha_v^{-1}
\end{tikzcd}$$

and the ordering is unique up to the action of the Weyl group $D_8 = \mathrm{Sym}(\square)$. In particular, the data of the quadruple $[\alpha_v, \beta_v, \gamma_v\beta_v^{-1}, \gamma_v\alpha_v^{-1}]$ carries with it the value of the similitude character.

We will be concerned with the case that the central character of π is given by $a \mapsto |a|^2$, in which case the Hecke parameters have the form $[\alpha_v, \beta_v, q_v\beta_v^{-1}, q_v\alpha_v^{-1}]$.

2.4.10. *Iwahori Hecke operators.*

Definition 2.4.11. We say that an unramified principal series $\pi(\chi)$ is *general* if the Hecke parameters are pairwise distinct and no ratio of them is q_v. In particular, $\pi(\chi)$ is irreducible, and $|W \cdot \chi| = 8$.

We have Iwahori Hecke operators $U^{\mathrm{naive}}_{\mathrm{Iw}(v),i} = [\mathrm{Iw}(v)\beta_{v,i}\mathrm{Iw}(v)]$. The notation "$U^{\mathrm{naive}}$" is intended to indicated that we have not yet appropriately normalized these operators, as we will shortly do in the case that $v|p$. Then we have

Proposition 2.4.12. *Let π be a general unramified principal series with Hecke parameters $[\alpha_v, \beta_v, q_v\beta_v^{-1}, q_v\alpha_v^{-1}]$. Then $\pi^{\mathrm{Iw}(v)}$ is a direct sum of 8 one-dimensional simultaneous eigenspaces for the $U^{\mathrm{naive}}_{\mathrm{Iw}(v),i}$. For a given (ordered) choice of α_v and β_v the corresponding eigenvalues are $u_{v,0} = q_v^{-2}$, $u_{v,1} = \alpha_v$, and $u_{v,2} = q_v^{-1}\alpha_v\beta_v$.*

Proof. The first part is immediate from Propositions 2.4.3 and 2.4.4. To compute the eigenvalues, by the definition of the Hecke parameters and Proposition 2.4.6 we have $\alpha_v = q_v^{3/2}(\chi_1\chi_2\sigma)(\varpi_v)$, $\beta_v = q_v^{3/2}(\chi_1\sigma)(\varpi_v)$ and $q_v = q_v^3(\chi_1\chi_2\sigma^2)(\varpi_v)$. We then have $u_{v,i} = \delta_B(\beta_{v,i})^{-1/2}(\chi_1\otimes\chi_2\otimes\sigma)(\beta_{v,i})$, so that $u_{v,0} = (\chi_1\chi_2\sigma)^2(\varpi_v) = q_v^{-2}$, $u_{v,1} = q_v^{3/2}(\chi_1\chi_2\sigma)(\varpi_v) = \alpha_v$, $u_{v,2} = q_v^2(\chi_1^2\chi_2\sigma^2)(\varpi_v) = q_v^{-1}\alpha_v\beta_v$, as required. $\square$

Proposition 2.4.12 has the following converse:

Proposition 2.4.13. *Let π be an irreducible admissible representation of $\mathrm{GSp}_4(F_v)$, and suppose that $\pi^{\mathrm{Iw}(v)}$ contains an eigenvector for the $U^{\mathrm{naive}}_{\mathrm{Iw}(v),i}$ with eigenvalues u_{v_i} satisfying $u_{v_0} = q_v^{-2}$, $u_{v,1} = \alpha_v$ and $u_{v,2} = q_v^{-1}\alpha_v\beta_v$ such that no ratio of a pair of $[\alpha_v, \beta_v, q_v\beta_v^{-1}, q_v\alpha_v^{-1}]$ is q_v. Then π is the unramified principal series with Hecke parameters $[\alpha_v, \beta_v, q_v\beta_v^{-1}, q_v\alpha_v^{-1}]$.*

Proof. Reversing the calculation in the previous proof, we let $\chi = \chi_1 \otimes \chi_2 \otimes \sigma$ be the unramified character with $\chi_1(\varpi_v) = \alpha_v\beta_v q_v^{-1}$, $\chi_2(\varpi_v) = \alpha_v\beta_v^{-1}$, and $\sigma(\varpi_v) = \alpha_v^{-1}q_v^{-1/2}$. We see that there is an inequality $\mathrm{Hom}_{T(F_v)}(\pi^{\mathrm{Iw}(v)}, \chi) \neq \{0\}$, and hence $\mathrm{Hom}(\pi, \pi(\chi)) \neq \{0\}$ by Proposition 2.4.3 and Frobenius reciprocity. Finally, by Proposition 2.4.6, $\pi(\chi)$ is also irreducible. □

2.4.14. *Parahoric level Hecke operators for* GL_2. We will also need to consider certain parahoric Hecke algebra and investigate how they relate to the Iwahori Hecke algebra.

We begin by recalling some standard results for the group GL_2. We let $\mathrm{Iw}(v)' \subset \mathrm{GL}_2(\mathcal{O}_{F_v})$ be the Iwahori subgroup of matrices which are upper triangular modulo ϖ_v (we put a prime because $\mathrm{Iw}(v)$ is used to denote the Iwahori subgroup in $\mathrm{GSp}_4(\mathcal{O}_{F_v})$).

We introduce the following operators in the spherical Hecke algebra $\mathcal{H}_{\mathrm{Sph}}[1/p]$:

(1) $T^{\mathrm{GL}_2}_{v,1} = [\mathrm{GL}_2(\mathcal{O}_{F_v}) \begin{pmatrix} \varpi_v & 0 \\ 0 & 1 \end{pmatrix} \mathrm{GL}_2(\mathcal{O}_{F_v})]$,

(2) $T^{\mathrm{GL}_2}_{v,0} = [\mathrm{GL}_2(\mathcal{O}_{F_v}) \begin{pmatrix} \varpi_v & 0 \\ 0 & \varpi_v \end{pmatrix} \mathrm{GL}_2(\mathcal{O}_{F_v})]$.

We also define the following operators in the Iwahori Hecke algebra $\mathcal{H}_{\mathrm{Iw}(v)'}[1/p]$:

(1) $U^{\mathrm{GL}_2}_{v,1} = [\mathrm{Iw}(v)' \begin{pmatrix} \varpi_v & 0 \\ 0 & 1 \end{pmatrix} \mathrm{Iw}(v)']$,

(2) $U^{\mathrm{GL}_2}_{v,0} = [\mathrm{Iw}(v)' \begin{pmatrix} \varpi_v & 0 \\ 0 & \varpi_v \end{pmatrix} \mathrm{Iw}(v)']$,

(3) $e^{\mathrm{GL}_2}_{\mathrm{Sph}} = [\mathrm{GL}_2(\mathcal{O}_{F_v})]$.

For any element f of the centre of the Iwahori Hecke algebra, the element $e^{\mathrm{GL}_2}_{\mathrm{Sph}} f$ defines an element of the spherical Hecke algebra.

Lemma 2.4.15. *The centre $Z(\mathcal{H}_{\mathrm{Iw}(v)'}[1/p])$ of the Iwahori Hecke algebra is generated by $U^{\mathrm{GL}_2}_{v,0}$ and $q_v U^{\mathrm{GL}_2}_{v,0}(U^{\mathrm{GL}_2}_{v,1})^{-1} + U^{\mathrm{GL}_2}_{v,1}$, the map $e^{\mathrm{GL}_2}_{\mathrm{Sph}} : Z(\mathcal{H}_{\mathrm{Iw}(v)'}[1/p]) \to \mathcal{H}_{\mathrm{Sph}}[1/p]$ is an isomorphism and we have the following identities:*

(1) $e^{\mathrm{GL}_2}_{\mathrm{Sph}} U^{\mathrm{GL}_2}_{v,0} = T^{\mathrm{GL}_2}_{v,0}$,

(2) $e^{\mathrm{GL}_2}_{\mathrm{Sph}}(q_v U^{\mathrm{GL}_2}_{v,0}(U^{\mathrm{GL}_2}_{v,1})^{-1} + U^{\mathrm{GL}_2}_{v,1}) = T_{v,1}$.

Proof. This follows from [HKP10, §1, §2, §4.6]. □

2.4.16. *Klingen level Hecke operators.* We have Klingen Hecke operators $U^{\mathrm{naive}}_{\mathrm{Kli}(v),i} = [\mathrm{Kli}(v)\beta_{v,i}\mathrm{Kli}(v)]$.

Proposition 2.4.17. *Let π be a general unramified principal series with Hecke parameters $[\alpha_v, \beta_v, q_v\beta_v^{-1}, q_v\alpha_v^{-1}]$. Then $\pi^{\mathrm{Kli}(v)}$ is a direct sum of 4 one-dimensional simultaneous eigenspaces for the $U^{\mathrm{naive}}_{\mathrm{Kli}(v),i}$. For a given choice of $\{\alpha_v, \beta_v\}$, the eigenvalues are $u_{v,0} = q_v^{-2}$, $u_{v,1} = \alpha_v + \beta_v$, and $u_{v,2} = q_v^{-1}\alpha_v\beta_v$.*

Proof. This follows from a direct computation, see [GT05, Prop. 3.2.1, Cor. 3.2.2]. □

Remark 2.4.18. We sketch another (related) proof of Proposition 2.4.17. Let us denote by $\mathcal{H}_{\mathrm{Iw}(v)}[1/p]$ the Iwahori Hecke algebra and by $Z_{\mathrm{Kli}(v)}(\mathcal{H}_{\mathrm{Iw}(v)}[1/p])$ the subalgebra generated by $U^{\mathrm{naive}}_{\mathrm{Iw}(v),1}+q_v(U^{\mathrm{naive}}_{\mathrm{Iw}(v),1})^{-1}U^{\mathrm{naive}}_{\mathrm{Iw}(v),2}, U^{\mathrm{naive}}_{\mathrm{Iw}(v),2}, U^{\mathrm{naive}}_{\mathrm{Iw}(v),0}$. One checks that $Z_{\mathrm{Kli}(v)}(\mathcal{H}_{\mathrm{Iw}(v)}[1/p])$ commutes with $e_{\mathrm{Kli}(v)}=[\mathrm{Kli}(v)]$ by using Bernstein's relation ([HKP10, §1.15]). Therefore we get a map: $e_{\mathrm{Kli}(v)}: Z_{\mathrm{Kli}(v)}(\mathcal{H}_{\mathrm{Iw}(v)}[1/p])\to \mathcal{H}_{\mathrm{Kli}(v)}[1/p]$ where $\mathcal{H}_{\mathrm{Kli}(v)}[1/p]$ is the Klingen Hecke algebra. We claim that:

- $e_{\mathrm{Kli}(v)}(U^{\mathrm{naive}}_{\mathrm{Iw}(v),1}+q_v(U^{\mathrm{naive}}_{\mathrm{Iw}(v),1})^{-1}U^{\mathrm{naive}}_{\mathrm{Iw}(v),2})=U^{\mathrm{naive}}_{\mathrm{Kli}(v),1}$,
- $e_{\mathrm{Kli}(v)}U^{\mathrm{naive}}_{\mathrm{Iw}(v),2}=U^{\mathrm{naive}}_{\mathrm{Kli}(v),2}$,
- $e_{\mathrm{Kli}(v)}U^{\mathrm{naive}}_{\mathrm{Iw}(v),0}=U^{\mathrm{naive}}_{\mathrm{Kli}(v),0}$.

The claim can be checked after restricting all these functions to the Levi $\mathrm{GL}_2\times\mathrm{GL}_1$ of the Klingen parabolic by [Vig98, Prop. II.5], so it follows from Lemma 2.4.15. The result then follows from Proposition 2.4.12.

Remark 2.4.19. Proposition 2.4.17 could also be proved using Jacquet modules (as could analogous results for invariants at other level structures which admit parahoric factorizations).

Proposition 2.4.20. *Let π be an irreducible admissible representation of $\mathrm{GSp}_4(F_v)$, and suppose that $\pi^{\mathrm{Kli}(v)}$ contains an eigenvector for the $U^{\mathrm{naive}}_{\mathrm{Kli}(v),i}$ with eigenvalues u_{v_i} satisfying $u_{v_0}=q_v^{-2}$, $u_{v,1}=\alpha_v+\beta_v$ and $u_{v,2}=q_v^{-1}\alpha_v\beta_v$ such that no ratio of a pair of $\{\alpha_v,\beta_v,q_v\beta_v^{-1},q_v\alpha_v^{-1}\}$ is q_v. Then π is the unramified principal series with Hecke parameters $[\alpha_v,\beta_v,q_v\beta_v^{-1},q_v\alpha_v^{-1}]$.*

Proof. As in the proof of Proposition 2.4.13, we deduce from $\pi^{\mathrm{Iw}(v)}\neq 0$ that π is a constituent of an unramified principal series representation. The central character of such a constituent is unramified and so is determined by the value on Frobenius. From the equation $u_{v_0}=q_v^{-2}$, we deduce that the central character of π is $|\cdot|^2$, and hence the central character of $\pi\otimes|\nu|^{-3/2}$ is $|\cdot|^{-1}$, and hence that the similitude character of $\mathrm{rec}_{\mathrm{GT},p}(\pi\otimes|\nu|^{-3/2})$ is the inverse of the cyclotomic character ε^{-1}. In particular, the value of the similitude character of the Weil–Deligne representation on Frob_v is q_v, and thus π is a constituent of an unramified principal series representation with Hecke parameters $[\alpha'_v,\beta'_v,q_v(\beta'_v)^{-1},q_v(\alpha'_v)^{-1}]$. (Note that the ordering of these eigenvalues above is determined up to the action of D_8.) Comparing to Proposition 2.4.17, without loss of generality, we may rearrange the Hecke parameters of π so that we deduce the two equations

$$\alpha'_v+\beta'_v=\alpha_v+\beta_v,\qquad \alpha'_v\beta'_v=\alpha_v\beta_v,$$

and thus (again up to reordering) $\alpha'_v=\alpha_v$ and $\beta'_v=\beta_v$. By Proposition 2.4.6, the principal series π is irreducible. □

2.4.21. *Generic unipotent representations.* We say that a $\mathrm{GSp}_4(\overline{E})$-valued Weil–Deligne representation r is *generic* if $\mathrm{ad}(r)(1)$ has no invariants, and is *unipotent* if r^{ss} is unramified.

Proposition 2.4.22. *Let r be unipotent. Then the L-packet corresponding to r contains a generic representation if and only if r is generic.*

Proof. By the main theorem of [GT11a] (part vii), the L-packet $L(r)$ contains a generic representation if and only if the adjoint L-factor $L(s, \mathrm{ad}(r^{F-\mathrm{ss}}))$ is holomorphic at $s = 1$, which, by definition, is easily seen to be equivalent to the statement that $\mathrm{ad}(r^{F-\mathrm{ss}})(1)$ has no invariants. Thus we are reduced to checking that r is generic if and only if $r^{F-\mathrm{ss}}$ is generic. Let W denote the vector space underlying the representation $\mathrm{ad}(r)(1)$. We are reduced to showing that $\mathrm{Hom}(\overline{E}, W) = 0$ if and only if $\mathrm{Hom}(\overline{E}, W^{F-\mathrm{ss}}) = 0$.

One implication is trivial. For the reverse implication, a map from $\overline{E}$ to $W^{F-\mathrm{ss}}$ is the same as giving a vector x in W which lies in the kernel of N and is a generalized eigenvector for the Frobenius ϕ with eigenvalue 1. For a suitable choice of $n \in \mathbf{N}$, the vector $y = (\phi - 1)^n x$ will be non-zero and a genuine eigenvector for ϕ with eigenvalue one. On the other hand, since x lies in the kernel of N, so does ϕx, because $N\phi x = q_v^{-1}\phi N x = 0$. Similarly, any polynomial in ϕ applied to x also lies in the kernel of N. Thus y also lies in the kernel of N and gives rise to a nonzero element of $\mathrm{Hom}(\overline{E}, W)$. □

2.4.23. *Normalized Hecke operators, ordinary representations, and ordinary projectors.* In this section, we assume that $v|p$. We fix integers $k \geq l \geq 2$, and $k \equiv l \pmod 2$ (these will correspond to the weights of our automorphic forms; see Section 2.6). Then we will consider normalized Hecke operators at Iwahori and Klingen level defined by

$$U_{\mathrm{Iw}(v),0} = p^2 U^{\mathrm{naive}}_{\mathrm{Iw}(v),0} \qquad U_{\mathrm{Kli}(v),0} = p^2 U^{\mathrm{naive}}_{\mathrm{Kli}(v),0}$$
$$U_{\mathrm{Iw}(v),1} = p^{(k+l)/2-2} U^{\mathrm{naive}}_{\mathrm{Iw}(v),1} \qquad U_{\mathrm{Kli}(v),1} = p^{(k+l)/2-2} U^{\mathrm{naive}}_{\mathrm{Kli}(v),1}$$
$$U_{\mathrm{Iw}(v),2} = p^{k-1} U^{\mathrm{naive}}_{\mathrm{Iw}(v),2} \qquad U_{\mathrm{Kli}(v),2} = p^{k-1} U^{\mathrm{naive}}_{\mathrm{Kli}(v),2}$$

We will often write $U_{v,i}$ for the operators $U_{\mathrm{Iw}(v),i}$ when the context is clear. We will also keep writing $U_{v,0}$ for the Hecke operator $p^2[K_v\beta_{v,0}K_v]$ for any subgroup K_v of $\mathrm{Iw}(v)$ (because $U_{v,0}$ lies in the centre of the Iwahori Hecke algebra and therefore $p^2[K_v\beta_{v,0}K_v] = e_{K_v}U_{v,0}$). We will also often write $U_{v,2}$ for the Hecke operator $U_{\mathrm{Kli}(v),2}$ for the same reason (see Remark 2.4.18). We can and do also normalize the Siegel Hecke operators in the same way, so that for example $U_{\mathrm{Si}(v),1} = p^{(k+l)/2-2}U^{\mathrm{naive}}_{\mathrm{Si}(v),1}$.

An irreducible smooth $\overline{E}[\mathrm{GSp}_4(F_v)]$-representation with central character $|\cdot|^2$ is said to be *ordinary* of weights $k \geq l \geq 2$ if there exists an eigenvector $v \in \pi^{\mathrm{Iw}(v)}$ for $U_{\mathrm{Iw}(v),i}$ with eigenvalues $u_{v,i}$ with $v_p(u_{v,i}) = 0$. If α_v and β_v are defined by $\alpha_v = u_{v,1}$, $\beta_v = u_{v,2}/u_{v,1}$, then, by Proposition 2.4.4, π is a constituent of an unramified principal series with Hecke parameters

$$[\alpha_v p^{2-(k+l)/2}, \beta_v p^{-(k-l)/2}, \beta_v^{-1} p^{1+(k-l)/2}, \alpha_v^{-1} p^{(k+l)/2-1}].$$

We say that π is *p-distinguished* if these four Satake parameters are pairwise distinct, or in other words if either $l > 2$ or $\alpha_v \neq \beta_v$.

If $l > 2$, then again by Proposition 2.4.4, $v \in \pi^{\mathrm{Iw}(v)}$ is the unique eigenvector (up to scale) with unit eigenvalues for the $U_{\mathrm{Iw}(v),i}$. In this case, the ordered pair (α_v, β_v) is uniquely determined by π, and we call (α_v, β_v) the *ordinary Hecke parameters* of π. If $l = 2$ and π is p-distinguished, then there may also be an eigenvector $v' \in \pi^{\mathrm{Iw}(v)}$ with unit eigenvalues $U_{\mathrm{Iw}(v),1}v' = \beta_v v'$, $U_{\mathrm{Iw}(v),2}v' = \alpha_v\beta_v v'$ (we will see below that in fact such a v' always exists.) Thus at least the set $\{\alpha_v, \beta_v\}$ is determined by π and we again call them the *ordinary Hecke parameters* of π.

We let e_{reg} be the ordinary projector (in the sense of Section 2.11) associated to $U_{\mathrm{Iw}(v),1}U_{\mathrm{Iw}(v),2}$, and let e_{irreg} be the ordinary projector associated to $U_{\mathrm{Kli}(v),2}$.

Proposition 2.4.24. *Let π be an ordinary p-distinguished representation of weights $k \geq l \geq 2$, with ordinary Hecke parameters (α_v, β_v) (or $\{\alpha_v, \beta_v\}$ if $l = 2$). Assume that either $k > l > 3$ or $l = 2$.*

(1) If $k > l > 3$ or if $l = 2$ and $k > 2$, then π is an irreducible principal series.
(2) If $k = l = 2$, then in the sense of the tables of [RS07c, §1]*, π is a representation of type* Va *if $\{\alpha_v, \beta_v\} = \{1, -1\}$,* IIIa *if $\alpha_v\beta_v = 1$,* IIa *if $\#\{\alpha_v, \beta_v\} \cap \{1, -1\} = 1$, or otherwise is an irreducible unramified principal series.*

In all cases, π is generic and the L-packet $L(\pi)$ of π contains no other ordinary representations. Moreover:

(1) $k > l > 3$ then
$$\dim e_{reg}\pi^{\mathrm{Iw}(v)} = 1$$
on which $U_{\mathrm{Iw}(v),i}$ has eigenvalues $1, \alpha_v, \alpha_v\beta_v$ for $i = 0, 1, 2$.
(2) If $l = 2$ then
$$\dim e_{reg}\pi^{\mathrm{Iw}(v)} = 2$$
and there are two eigenspaces for $U_{\mathrm{Iw}(v),i}$, with eigenvalues $1, \alpha_v, \alpha_v\beta_v$ and $1, \beta_v, \alpha_v\beta_v$ respectively, and moreover
$$\dim e_{irreg}\pi^{\mathrm{Kli}(v)} = 1$$
with $U_{\mathrm{Kli}(v),i}$ eigenvalues $1, \alpha_v + \beta_v, \alpha_v\beta_v$, for $i = 0, 1, 2$.

Proof. As remarked above, by Proposition 2.4.4, π is a constituent of an unramified principal series with Hecke parameters
$$[\alpha_v p^{2-(k+l)/2}, \beta_v p^{-(k-l)/2}, \beta_v^{-1} p^{1+(k-l)/2}, \alpha_v^{-1} p^{(k+l)/2-1}].$$
If either $k > l > 3$ or $l = 2$ and $k > 2$, no ratio of a pair of these parameters can be p, and hence π is an irreducible principal series by Proposition 2.4.6.

In the remaining case, $k = l = 2$, the Satake parameters are $[\alpha_v, \beta_v, \beta_v^{-1}p, \alpha_v^{-1}p]$, and the corresponding principal series may be reducible when one of $\alpha_v^2, \beta_v^2, \alpha_v\beta_v$ is equal to 1. The constituents of these principal series are listed in the tables [RS07c, §1]. The case that either $\alpha_v^2 = 1$ or $\beta_v^2 = 1$ but not both corresponds to type II, the case that $\alpha_v^2 = \beta_v^2 = 1$ corresponds to type V, and the case that $\alpha_v\beta_v = 1$ corresponds to type III.

For each constituent π of such a principal series, the tables give a computation of the Jacquet module π_U^{ss}, which is equal to π_U because $\alpha_v \neq \beta_v$. This allows us, by Proposition 2.4.3, to determine the simultaneous eigenvalues of the $U_{\mathrm{Iw}(v),i}$ on $\pi^{\mathrm{Iw}(v)}$. At this point the result follows from an inspection of the tables. □

We now turn to the global situation. Recall that we have fixed an isomorphism $\imath : \overline{\mathbf{Q}}_p \cong \mathbf{C}$, so that in particular in the following definition we can and do identify the infinite places of F with the places dividing p. See Section 2.6 for our conventions regarding the weights of automorphic representations.

Definition 2.4.25. Let π be a cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$ with central character $|\cdot|^2$ and weight $(k_v, l_v)_{v|\infty}$, where $k_v \geq l_v \geq 2$ and $k_v \equiv l_v \pmod 2$ for all $v|\infty$. Then we say that π is *ordinary* if for each place $v|p$, π_v is ordinary of weights $k_v \geq l_v \geq 2$.

The following proposition will be useful for going between ordinary p-adic modular forms and ordinary automorphic representations. For each subset $I \subset S_p$ we set

$$K_p(I) = \prod_{v \in I} \mathrm{Kli}(v) \prod_{v \in I^c} \mathrm{Iw}(v).$$

We also let $e(I) = \prod_{v \in I} e_{\mathrm{irreg}} \prod_{v \notin I} e_{\mathrm{reg}}$.

Proposition 2.4.26. *Let π be a cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$ of weight $(k_v, l_v)_{v|\infty}$ with $k_v \geq l_v \geq 2$ and with central character $|\cdot|^2$, and fix tuples of p-adic units $(\alpha_v, \beta_v)_{v|p}$. Assume that for each $v \in S_p$, either $k_v > l_v > 3$ or $l_v = 2$ and $\alpha_v \neq \beta_v$.*

Let $I' = \{v \in S_p \mid l_v = 2\}$ and let $I \subset I'$ be a subset. Then π is ordinary with ordinary Hecke parameters $(\alpha_v, \beta_v)_{v|p}$ if and only if

$$(\otimes_{v \in S_p} \pi_v)^{K_p(I)}$$

contains a vector which is:

- *for each $v \in I^c$, an eigenvector for the normalized $U_{\mathrm{Iw}(v),0}$, $U_{\mathrm{Iw}(v),1}$, and $U_{\mathrm{Iw}(v),2}$, with respective eigenvalues $1, \alpha_v$, and $\alpha_v\beta_v$, and*
- *for each $v \in I$, an eigenvector for $U_{\mathrm{Kli}(v),0}$, $U_{\mathrm{Kli}(v),1}$, and $U_{\mathrm{Kli}(v),2}$ with respective eigenvalues $1, \alpha_v + \beta_v$, and $\alpha_v\beta_v$.*

Moreover in this case

$$\dim e(I)(\otimes_{v \in S_p} \pi_v)^{K_p(I)} = 2^{\#(I'-I)}.$$

Note that if π is ordinary with ordinary Hecke parameters $(\alpha_v, \beta_v)_{v|p}$ but $v \notin I'$, then the $U_{\mathrm{Kli}(v),1}$ eigenvalue will *not* be of the form $\alpha_v + \beta_v$, but rather, up to some ordering of α_v and β_v, be of the form $\alpha_v + p^{l_v - 2}\beta_v$.

Proof. This is simply Proposition 2.4.24 applied for each $v \in S_p$. □

2.4.27. *An instance of the local Langlands correspondence.* Given a pair of characters $\chi_{v,1}, \chi_{v,2} : k(v)^\times \to \mathcal{O}^\times$, which we regard as characters of $\mathcal{O}_{F_v}^\times$ by inflation, we define a character of χ_v of $T(\mathcal{O})$ by

$$\chi_v : T(\mathcal{O}_{F_v}) \to \mathcal{O}^\times$$
$$(a, b, cb^{-1}, ca^{-1}) \mapsto \chi_{v,1}(ab^{-1})\chi_{v,2}(abc^{-1}).$$

Then if M is an $\mathcal{H}_1$-module, we write

$$M^{\chi_v} = \{m \in M \mid tm = \chi_v(t)m \ \forall t \in T(k(v))\}$$

and

$$M_{\chi_v} = M/\langle tm - \chi_v(t)m \mid t \in T(k(v)), m \in M\rangle.$$

Then we record:

Proposition 2.4.28. *If π is an irreducible smooth $\overline{E}[\mathrm{GSp}_4(F_v)]$-module with the property that $(\pi^{\mathrm{Iw}_1(v)})^{\chi_v} \neq \{0\}$, then, for all $\sigma \in W_{F_v}$,*

$$\det(X - \mathrm{rec}_{\mathrm{GT},p}(\pi)(\sigma)) = (X - \chi_{v,1}(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_{v,1}(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1})$$
$$(X - \chi_{v,2}(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_{v,2}(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}).$$

If, moreover, the characters $\chi_{v,1}, \chi_{v,1}^{-1}, \chi_{v,2}, \chi_{v,2}^{-1}$ are pairwise distinct, then there is an equality $\dim_{\overline{E}}(\pi^{\mathrm{Iw}_1(v)})^{\chi_v} = 1$.

Proof. This is an immediate consequence of Propositions 2.4.3, 2.4.4 and 2.4.6. □

2.4.29. *The case $q_v \equiv 1 \pmod p$.* We suppose from now on for the rest of this section that $q_v \equiv 1 \pmod p$. Recall that we have a homomorphism $T(F_v)/T(\mathcal{O}_{F_v})_1 \to \mathcal{H}_1^\times$, and thus an (injective) homomorphism $\mathcal{O}[T(F_v)/T(\mathcal{O}_{F_v})_1] \to \mathcal{H}_1$; we identify $\mathcal{O}[T(F_v)/T(\mathcal{O}_{F_v})_1]$ with its image in $\mathcal{H}_1$. Given elements $\overline{\alpha}_1, \overline{\alpha}_2 \in \overline{\mathbf{F}}_p^\times$, we let $\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}$ denote the kernel of the homomorphism $\mathcal{O}[T(F_v)/T(\mathcal{O}_{F_v})_1] \to \overline{\mathbf{F}}_p$ induced by the character $T(F_v)/T(\mathcal{O}_{F_v})_1 \to \overline{\mathbf{F}}_p^\times$ sending $T(\mathcal{O}_{F_v}) \mapsto 1$, $\mathrm{diag}(\varpi_v, \varpi_v, \varpi_v, \varpi) \mapsto 1$, $\mathrm{diag}(\varpi_v, \varpi_v, 1, 1) \mapsto \overline{\alpha}_1$, and $\mathrm{diag}(\varpi_v^2, \varpi_v, \varpi_v, 1) \mapsto \overline{\alpha}_1\overline{\alpha}_2$.

Proposition 2.4.30. *Let π be an irreducible smooth $\overline{E}[\mathrm{GSp}_4(F_v)]$-module with central character $|\cdot|^2$ and with $(\pi^{\mathrm{Iw}_1(v)})_{\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}} \neq \{0\}$. Suppose $\overline{\alpha}_1^{\pm 1}, \overline{\alpha}_2^{\pm 1}$ are pairwise distinct. Then*

$$\mathrm{rec}_{\mathrm{GT},p}(\pi) = \gamma_1 \oplus \gamma_2 \oplus \varepsilon^{-1}\gamma_2^{-1} \oplus \varepsilon^{-1}\gamma_1^{-1}$$

for characters γ_i of G_{F_v} with $\overline{\gamma}_i = \lambda_{\overline{\alpha}_i}$ (the unramified character taking Frob_v to $\overline{\alpha}_i$), and $T(k(v))$ acts on $(\pi^{\mathrm{Iw}_1(v)})_{\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}}$ via $(\gamma_i \circ \mathrm{Art}_{F_v})|_{\mathcal{O}_{F_v}^\times}$.

Proof. From Proposition 2.4.28, we know the characteristic polynomial of the corresponding representation, and thus immediately deduce that the semi-simplification of the Galois representation has the required form. It thus suffices to show that, under the hypothesis on $\overline{\alpha}_i$, that all Galois representations are semi-simple. Suppose otherwise. Two tamely ramified characters admit an extension if and only if their ratio is unramified and takes the value q_v on Frobenius. Since $q_v \equiv 1 \mod p$ and ε is trivial modulo p, this implies that $\overline{\alpha}_1^{\pm 1}, \overline{\alpha}_2^{\pm 1}$ are not distinct, a contradiction. □

Remark 2.4.31. Let Z be the centre of GSp_4, let Δ_v be the maximal p-power quotient of $T(k(v))/Z(k(v))$, and let $\Delta_v' = \ker(T(k(v)) \to \Delta_v)$. If the π of Proposition 2.4.30 additionally satisfies the condition that $(\pi^{\mathrm{Iw}_1(v)})^{\Delta_v'}_{\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}} \neq \{0\}$, then we immediately deduce that Δ_v also acts on $(\pi^{\mathrm{Iw}_1(v)})^{\Delta_v'}_{\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}}$ via $(\gamma_i \circ \mathrm{Art}_{F_v})|_{\mathcal{O}_{F_v}^\times}$.

We now prove some results about the Iwahori Hecke algebra (under our running assumption that $q_v \equiv 1 \pmod p$). We follow [KT17, §5] closely, and our proofs are essentially an immediate adaptation of their arguments from GL_n to GSp_4. As recalled above, we have an embedding $\mathcal{O}[X_*(T)] \hookrightarrow \mathcal{H}$. This can be refined to give the Bernstein presentation of $\mathcal{H}$ (see e.g. [HKP10, §1]), which is an algebra isomorphism

$$\mathcal{H} \cong \mathcal{O}[X_*(T)] \widetilde{\otimes}_{\mathcal{O}} \mathcal{O}[\mathrm{Iw}(v) \backslash \mathrm{GSp}_4(\mathcal{O}_{F_v})/\mathrm{Iw}(v)],$$

where the twisted tensor product $\widetilde{\otimes}_{\mathcal{O}}$ is determined by the following relations, where $s_\alpha \in W$ is simple, corresponding to the simple root α, and $\mu \in X_*(T)$:

$$T_{s_\alpha}\theta_\lambda = \theta_{s_\alpha(\lambda)}T_{s_\alpha} + \big(q_v - 1\big)\frac{\theta_{s(\lambda)} - \theta_\lambda}{1 - \theta_{-\alpha^\vee}}. \tag{2.4.32}$$

Here we are writing θ_μ for the image in $\mathcal{H}$ of the group element e_μ of $\mathcal{O}[X_*(T)]$ corresponding to μ, and for $w \in W$ we write $T_w := [\mathrm{Iw}(v)\dot{w}\mathrm{Iw}(v)]$ where $\dot{w} \in \mathrm{GSp}_4(\mathcal{O}_{F_v})$ is any representative for w.

Lemma 2.4.33. *There is a natural isomorphism $\mathcal{H} \otimes_{\mathcal{O}} k \cong k[X_*(T) \rtimes W]$.*

Proof. We claim that the natural k-linear map $k[W] \to k[\mathrm{Iw}(v) \backslash \mathrm{GSp}_4(\mathcal{O}_{F_v})/\mathrm{Iw}(v)]$ sending $w \mapsto T_w$ is an algebra isomorphism. Admitting this claim, note that since $q_v \equiv 1 \pmod p$, the relation (2.4.32) becomes

$$T_{s_\alpha}\theta_\lambda = \theta_{s_\alpha(\lambda)}T_{s_\alpha}$$

in $\mathcal{H} \otimes_{\mathcal{O}} k$, so that there is an isomorphism $k[X_*(T) \rtimes W] \to \mathcal{H} \otimes_{\mathcal{O}} k$ sending $e_\lambda w \mapsto \theta_\lambda T_w$, as required.

It remains to prove the claim. The Weyl group W is generated by s_1, s_2 with $s_1^2 = s_2^2 = (s_1 s_2)^4 = 1$, so it is enough to show that $k[\mathrm{Iw}(v)\backslash \mathrm{GSp}_4(\mathcal{O}_{F_v})/\mathrm{Iw}(v)]$ is generated by the elements T_{s_1}, T_{s_2}, subject to the same relations. This follows from the assumption that $q_v \equiv 1 \pmod p$; indeed, we have the usual relations $T_{s_i}^2 = (q_v - 1)T_{s_i} + q_v$ $(i = 1, 2)$, and $T_{s_1}T_{s_2}T_{s_1}T_{s_2} = T_{s_2}T_{s_1}T_{s_2}T_{s_1}$, which are easily seen to be equivalent to $T_{s_1}^2 = T_{s_2}^2 = (T_{s_1}T_{s_2})^4 = 1$, as required. □

Recall that by definition an $\mathcal{O}[\mathrm{GSp}_4(F_v)]$-module M is *smooth* if every element of M is fixed by some open compact subgroup of $\mathrm{GSp}_4(F_v)$, and it is *admissible* if it is smooth, and if for each open compact subgroup $U \subset \mathrm{GSp}_4(F_v)$, M^U is a finite $\mathcal{O}$-module.

Lemma 2.4.34. *If M is a smooth $\mathcal{O}[\mathrm{GSp}_4(F_v)]$-module, then the natural inclusion $M^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \subset M^{\mathrm{Iw}(v)}$ is canonically split by the Hecke operator*

$$\frac{1}{[\mathrm{GSp}_4(\mathcal{O}_{F_v}) : \mathrm{Iw}(v)]} e_{\mathrm{Sph}(v)}.$$

Proof. The Hecke operator $e_{\mathrm{Sph}(v)} \in \mathcal{H}$ induces the natural trace map $M^{\mathrm{Iw}(v)} \to M^{\mathrm{GSp}_4(\mathcal{O}_{F_v})}$, so that the composite map $M^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \to M^{\mathrm{Iw}(v)} \to M^{\mathrm{GSp}_4(\mathcal{O}_{F_v})}$ is given by multiplication by $[\mathrm{GSp}_4(\mathcal{O}_{F_v}) : \mathrm{Iw}(v)]$. Since $[\mathrm{GSp}_4(\mathcal{O}_{F_v}) : \mathrm{Iw}(v)] \equiv |W| = 8 \pmod p$ is a unit in $\mathcal{O}$, we are done. □

Corollary 2.4.35. *If M is a smooth $k[G]$-module, then $M^{\mathrm{Iw}(v)}$ is naturally a $k[W]$-module, and $M^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} = (M^{\mathrm{Iw}(v)})^W$.*

Proof. This is immediate from Lemmas 2.4.33 and 2.4.34. □

The centre of $\mathcal{H}$ is $\mathcal{O}[X_*(T)]^W$, and there is an isomorphism

$$\mathcal{O}[X_*(T)]^W \cong \mathcal{O}[\mathrm{GSp}_4(\mathcal{O}_{F_v})\backslash \mathrm{GSp}_4(F_v)/\mathrm{GSp}_4(\mathcal{O}_{F_v})]$$

given by $x \mapsto e_{\mathrm{Sph}(v)}x$ (where we are regarding x as an element of $\mathcal{H}$); this isomorphism agrees with the isomorphism given by the usual Satake isomorphism (see [HKP10, §4.6]). The classical description of $\mathcal{O}[X_*(T)]$ is as follows. Let x_0, x_1, and x_2 denote the following three cocharacters:

$$x_0 : t \to \mathrm{diag}(t, t, 1, 1),$$
$$x_1 : t \to \mathrm{diag}(1/t, 1, 1, t),$$
$$x_2 : t \to \mathrm{diag}(1, 1/t, t, 1).$$

Then $x_0^2 x_1 x_2$ is the cocharacter $t \mapsto \mathrm{diag}(t, t, t, t)$ and

$$\mathcal{O}[X_*(T)] = \mathcal{O}[x_0, x_1, x_2, (x_0^2 x_1 x_2)^{-1}] = \mathcal{O}[x_0, x_1, x_2, (x_0 x_1 x_2)^{-1}].$$

The effect of the involutions s_1, s_2, and $s_1 s_2 s_1 \in W$ on these cocharacters is to send (x_0, x_1, x_2) to

$$(x_0, x_2, x_1), (x_0 x_2, x_1, x_2^{-1}), (x_0 x_1, x_1^{-1}, x_2)$$

respectively. All of these involutions preserve $(x_0, x_0x_1, x_0x_2, x_0x_1x_2)$ considered as an unordered quadruple. Define elements $e_i(x_0, x_1, x_2) \in \mathcal{O}[X_*(T)]^W$, $0 \le i \le 4$, by the following formulae:

$$(X - x_0)(X - x_0x_1)(X - x_0x_2)(X - x_0x_1x_2) = \sum e_i(x_0, x_1, x_2) X^i.$$

The relation between the e_i and the Hecke operators $T_{v,i}$ is given by

$$\sum e_i(x_0,x_1,x_2)X^i = X^4 - q_v^{3/2}T_{v,1}X^3 + (q_v^2T_{v,2} + (1+q_v^2)T_{v,0})X^2 - q_v^{3/2}T_{v,0}T_{v,1}X + T_{v,0}^2.$$

Since we are assuming that $q_v \equiv 1 \pmod p$, and in our applications of these results in the global setting there is a twist which makes all of the powers of q_v integral (as in (2.4.8)), we will ignore all powers of $q_v^{1/2}$ from now on.

Given any triple $\gamma := (\gamma_0,\gamma_1,\gamma_2)$ and $w \in W$, let $((w\gamma)_0,(w\gamma)_1,(w\gamma)_2)$ denote the triple obtained by substituting in γ_i for x_i in the action of W on $\mathcal{O}[X_*(T)]$ described above.

Lemma 2.4.36. *Let M be an $\mathcal{H}\otimes_{\mathcal{O}} k$-module which is finite-dimensional over k. Suppose that $e_{\mathrm{Sph}(v)}M \neq 0$, and that there is a triple γ_0, γ_1, γ_2 with $\gamma_0^2\gamma_1\gamma_2 = 1$ such that $(\gamma_1-1)(\gamma_2-1)(\gamma_1-\gamma_2)(\gamma_1\gamma_2-1) \neq 0$; equivalently, writing $\alpha_1 = \gamma_0$, $\alpha_2 = \gamma_0\gamma_1$, suppose that*

$$\alpha_1, \alpha_2, 1/\alpha_2, 1/\alpha_1$$

are pairwise distinct. Suppose also that the following operators act by zero on the module $e_{\mathrm{Sph}(v)}M$:

$$T_{v,0} - 1\ T_{v,1} - e_1(\gamma_0,\gamma_1,\gamma_2), T_{v,2} + 2T_{v,0} - e_2(\gamma_0,\gamma_1,\gamma_2).$$

Then, for each $w \in W$, the maximal ideal

$$\mathfrak{m}_w = (x_0 - (w\gamma)_0, x_1 - (w\gamma)_1, x_2 - (w\gamma)_2) \subset k[X_*(T)]$$

is in the support of M.

Proof. Let $\mathfrak{n} \subset k[X_*(T)]^W$ be the ideal

$$\mathfrak{n} = (e_1(x_0,x_1,x_2) - e_1(\gamma_0,\gamma_1,\gamma_2), \ldots, e_4(x_0,x_1,x_2) - e_4(\gamma_0,\gamma_1,\gamma_2), x_0^2x_1x_2 - \gamma_0^2\gamma_1\gamma_2).$$

Then, by assumption, we have $e_{\mathrm{Sph}(v)}M \subset M[\mathfrak{n}]$, so that in particular $M_{\mathfrak{n}} \neq 0$. The assumptions on γ_i imply that all the ideals $\mathfrak{m}_w$ are distinct. We may view $\mathfrak{n}$ as an ideal in $k[X_*(T)]$. The support of $\mathfrak{n}$ in $k[X_*(T)]$ corresponds to triples $(\gamma_0,\gamma_1,\gamma_2)$ (or equivalently, pairs (α_1,α_2)) such that α_1, α_2, α_2^{-1}, and α_1^{-1} are roots of the polynomial $\sum e_i(\gamma_0,x_1,x_2)X^i$. Hence the support of $\mathfrak{n} \subset k[X_*(T)]$ consists exactly of the maximal ideals $\mathfrak{m}_w$, and the product of the $\mathfrak{m}_w$ is precisely the radical of $\mathfrak{n}$. The ring $k[X_*(T)]_{\mathfrak{n}}$ is thus a semi-local ring which is isomorphic to $\oplus_{w\in W} k[X_*(T)]_{\mathfrak{m}_w}$, and correspondingly we may write $M_{\mathfrak{n}} = \oplus_{w\in W} M_{\mathfrak{m}_w}$. It follows that $M_{\mathfrak{m}_w} \neq 0$ for at least one $w \in W$. Considering the action of W on the set of maximal ideals of $k[X_*(T)]$ in the support of M, we see that in fact $M_{\mathfrak{m}_w} \neq 0$ for all $w \in W$, as required. □

Lemma 2.4.37. *Let M be an $\mathcal{H}\otimes_{\mathcal{O}} k$-module which is finite-dimensional over k. Suppose that for each maximal ideal $\mathfrak{n} \subset k[X_*(T)]^W$ in the support of M, the degree four polynomial*

$$\sum e_i(x_0,x_1,x_2)X^i \in k[X_*(T)]^W[X]$$

has roots $(\gamma_0,\gamma_0\gamma_1,\gamma_0\gamma_2,\gamma_0\gamma_1\gamma_2)$ modulo $\mathfrak{n}$ satisfying $(\gamma_1-1)(\gamma_2-1)(\gamma_1-\gamma_2)(\gamma_1\gamma_2-1) \neq 0$ and $\gamma_0^2\gamma_1\gamma_2 = 1$. Equivalently, writing $\gamma_0 = \alpha_1$, $\gamma_0\gamma_1 = \alpha_2$, assume that $\gamma_0^2\gamma_1\gamma_2 = 1$ and that

$$\alpha_1, \alpha_2, 1/\alpha_2, 1/\alpha_1$$

are pairwise distinct. Then $e_{\mathrm{Sph}(v)}M \neq 0$. *If, furthermore, there is a unique maximal ideal* $\mathfrak{n} \subset k[X_*(T)]^W$ *in the support of* M, *then for each maximal ideal* $\mathfrak{m} \subset k[X_*(T)]$ *in the support of* M, *the maps*

$$k[W] \otimes_k M_{\mathfrak{m}} \to M,$$

$$w \otimes x \mapsto w \cdot x,$$

and

$$M_{\mathfrak{m}} \to e_{\mathrm{Sph}(v)}M,$$

$$x \mapsto e_{\mathrm{Sph}(v)} \cdot x$$

are both isomorphisms.

Proof. After possibly enlarging k, we can and do assume that the γ_i arising from the roots of the degree four polynomial above lie in k. As in the proof of Lemma 2.4.36, there exist $|W| = 8$ distinct ideals $\mathfrak{m}_w$ such that $M_{\mathfrak{n}} \simeq \oplus_{w \in W} M_{\mathfrak{m}_w}$, where $\mathfrak{m} = (t_0 - \gamma_0, t_1 - \gamma_1, t_2 - \gamma_2) \subset k[X_*(T)]$. Since $M_{\mathfrak{n}} \neq 0$, we may assume that $M_{\mathfrak{m}_w} \neq 0$ for some and hence all $\mathfrak{m}_w$. The operator $e_{\mathrm{Sph}(v)}$ acts by averaging over the action of the Weyl group. It follows (because the $\mathfrak{m}_w$ are distinct) that the map $e_{\mathrm{Sph}(v)} : M_{\mathfrak{m}} \to \oplus_{w \in W} M_{\mathfrak{m}_w} = M_{\mathfrak{n}}$ is an injection, and thus $e_{\mathrm{Sph}(v)}M \neq 0$.

Suppose that $\mathfrak{n}$ is the only maximal ideal of $k[X_*(T)]$ in the support of M. Then the maximal ideals of $k[X_*(T)]$ in the support of M are necessarily of the form $\mathfrak{m}_w$, and we have $M = \oplus_{w \in W} M_{\mathfrak{m}_w} = \oplus_{w \in W} w \cdot M_{\mathfrak{m}}$, and the rest of the lemma follows immediately. □

Remark 2.4.38. Note that (using as usual that $q_v = 1$ in k) we have that $U_{v,0}^{\mathrm{naive}} = x_0^2 x_1 x_2$, and if this equals 1, then $U_{v,1}^{\mathrm{naive}} = x_0$ and $U_{v,2}^{\mathrm{naive}} = (s_1 s_2 s_1) x_1$. Consequently we see for example that if the hypotheses of Lemma 2.4.36 hold then $(U_{v,0} - 1, U_{v,1} - \alpha_1, U_{v,2} - \alpha_1\alpha_2)$ is in the support of M.

2.5. **Purity.** Let K be a finite extension of $\mathbf{Q}_p$ for some p, with residue field of order q. Following [TY07, §1], we say that a Weil–Deligne representation (W, r, N) of W_K on a vector space W over an algebraically closed field Ω which is of characteristic 0 and of the same cardinality as $\mathbf{C}$ is *pure of weight* w if there is an exhaustive and separated ascending filtration Fil_i of W such that

- each $\mathrm{Fil}_i W$ is invariant under r;
- if $\sigma \in W_K$ maps to $\mathrm{Frob}_K^{v(\sigma)}$, then all eigenvalues of $r(\sigma)$ on $\mathrm{gr}_i W$ are Weil $q^{iv(\sigma)}$-numbers;
- and for all j we have $N^j : \mathrm{gr}_{w+j} W \xrightarrow{\sim} \mathrm{gr}_{w-j} W$. (Note that necessarily we have $N \mathrm{Fil}_i W \subset \mathrm{Fil}_{i-2} W$.)

Recall that for a Weil–Deligne representation (r, N), we defined in Section 2.3 $n(r, N)$ to be the rank of N.

Lemma 2.5.1. *If* (V, r) *is a semisimple representation of* W_K, *then there is at most one choice of* N *for which* (V, r, N) *is a pure Weil–Deligne representation. If such an* N *exists, then the corresponding Weil–Deligne representation is the unique choice which maximizes* $n(r, N)$.

Proof. The uniqueness of N is [TY07, Lem. 1.4(4)]. The maximality follows easily, using that by definition all of the induced maps $N^j : \mathrm{gr}_{w+j} W \to \mathrm{gr}_{w-j} W$ are isomorphisms if and only if (V, r, N) is pure. □

2.6. **Archimedean L-parameters.** We now recall some notation for archimedean L-parameters following [Mok14, §3.1] (although our w has the opposite sign to this reference). Recall that $W_{\mathbf{R}} = \mathbf{C}^\times \cup \mathbf{C}^\times j$, where $jzj^{-1} = \overline{z}$ and $j^2 = -1$. Let $w \in \mathbf{R}$. For an integer $n \geq 0$, let $\phi_{w,n} : W_{\mathbf{R}} \to \mathrm{GL}_2(\mathbf{C})$ be the L-parameter given by

$$z \mapsto |z|^w \cdot \begin{pmatrix} (z/\overline{z})^{n/2} & \\ & (z/\overline{z})^{-n/2} \end{pmatrix} = |z|^w \begin{pmatrix} z^n|z|^{-n} & \\ & z^{-n}|z|^n \end{pmatrix}$$

and

$$j \mapsto \begin{pmatrix} & 1 \\ (-1)^n & \end{pmatrix}.$$

The determinant of $\phi_{w,n}$ is equal to $|z|^{2w}$ if n is odd and $\mathrm{sgn} \cdot |z|^{2w}$ if n is even, where $\mathrm{sgn} : W_{\mathbf{R}} \to \mathbf{C}^\times$ is the degree two character which is -1 on j (and trivial on $\mathbf{C}^\times$). We also write $\phi_{w,n}$ for the restriction of $\phi_{w,n}$ to $W_{\mathbf{C}}$. The $\mathrm{GL}_2(\mathbf{R})$ and $\mathrm{GL}_2(\mathbf{C})$ representations corresponding to the L-parameter $\phi_{w,n}$ are cohomological if and only if $n > 0$ and $w \in \mathbf{Z}$ satisfies $w + n \equiv 1 \mod 2$.

Let $m_1 > m_2 \geq 0$ be integers, and let $w \in \mathbf{R}$. Then we write $\phi_{(w;m_1,m_2)} : W_{\mathbf{R}} \to \mathrm{GSp}_4(\mathbf{C})$ for the L-parameter sending

$$z \mapsto |z|^w \cdot \begin{pmatrix} (z/\overline{z})^{(m_1+m_2)/2} & & & \\ & (z/\overline{z})^{(m_1-m_2)/2} & & \\ & & (z/\overline{z})^{-(m_1-m_2)/2} & \\ & & & (z/\overline{z})^{-(m_1+m_2)/2} \end{pmatrix}$$

and

$$j \mapsto \begin{pmatrix} & & & 1 \\ & & 1 & \\ & (-1)^{m_1+m_2} & & \\ (-1)^{m_1+m_2} & & & \end{pmatrix}.$$

Note that $\phi_{(w;m_1,m_2)}$ is viewed as having image in $\mathrm{GSp}_4(\mathbf{C})$ with respect to our particular choice of model for $\mathrm{GSp}_4(\mathbf{C})$ where J is anti-diagonal. In particular, the image of j under the composite of $\phi_{(w;m_1,m_2)}$ with the similitude character is $(-1)^{m_1+m_2}$. With respect to the explicit inclusion of

$$\{(A, B) \subset \mathrm{GL}_2(\mathbf{C}) \times \mathrm{GL}_2(\mathbf{C}) \mid \det(A) = \det(B)\} \subset \mathrm{GSp}_4(\mathbf{C})$$

given in §2.2, we immediately observe that the composite of $\phi_{(w;m_1,m_2)}$ with the inclusion $\mathrm{GSp}_4(\mathbf{C}) \to \mathrm{GL}_4(\mathbf{C})$ identifies $\phi_{(w;m_1,m_2)}$ with $\phi_{w,m_1+m_2} \oplus \phi_{w,m_1-m_2}$ (note that $(-1)^{m_1+m_2} = (-1)^{m_1-m_2}$). The L-packet of $\mathrm{GSp}_4(\mathbf{R})$ corresponding to $\phi_{(w;m_1,m_2)}$ consists of two elements $\pi^H_{(w;m_1,m_2)}$ and $\pi^W_{(w;m_1,m_2)}$. When $m_2 = 0$, they are (up to twist) non-degenerate limits of discrete series, and when $m_2 > 0$, they are (up to twist) discrete series. The representations $\pi^H_{(w;m_1,m_2)}$ and $\pi^W_{(w;m_1,m_2)}$ are respectively holomorphic and generic. Their central character is given by $a \mapsto a^w$, and they are tempered when $w = 0$. The minimal K-type of $\pi^H_{(w;m_1,m_2)}$ is the representation $\det^{m_2+2} \otimes \mathrm{Sym}^{m_1-m_2-1} \mathbf{C}^2$ of $U(2)$. (See for example [Sch17] for these facts and their proofs.)

Lemma 2.6.1 (Inductions of real archimedean parameters to $\mathrm{GL}_4(\mathbf{C})$)**.**

(1) The induction $\mathrm{Ind}_{W_{\mathbf{C}}}^{W_{\mathbf{R}}} \phi_{w,n} : W_{\mathbf{R}} \to \mathrm{GL}_4(\mathbf{C})$ *is conjugate to* $\phi_{w,n} \oplus \phi_{w,n}$.

(2) The composite map

$$\phi_{(w;n,0)} : W_{\mathbf{R}} \to \mathrm{GSp}_4(\mathbf{C}) \to \mathrm{GL}_4(\mathbf{C})$$

is conjugate to $\phi_{w,n} \oplus \phi_{w,n}$.

(3) *If* $\varphi : W_{\mathbf{C}} \to \mathrm{GL}_2(\mathbf{C})$ *is such that* $\mathrm{Ind}_{W_{\mathbf{C}}}^{W_{\mathbf{R}}} \varphi$ *is conjugate to* $\phi_{w,n} \oplus \phi_{w,n}$, *and* $n \neq 0$, *then either* $\varphi \cong \phi_{w,n}$, *or* φ *is one of the scalar* L*-parameters sending* z *to one of*

$$|z|^w \cdot \begin{pmatrix} z^n|z|^{-n} & 0 \\ 0 & z^n|z|^{-n} \end{pmatrix} \text{ or } |z|^w \cdot \begin{pmatrix} z^{-n}|z|^n & 0 \\ 0 & z^{-n}|z|^n \end{pmatrix}.$$

(4) *If* $\varphi, \varphi' : W_{\mathbf{R}} \to \mathrm{GL}_2(\mathbf{C})$ *are such that* $\varphi \oplus \varphi'$ *is conjugate to* $\phi_{w,n} \oplus \phi_{w,n}$, *and* $n \neq 0$, *then* $\varphi \cong \varphi' \cong \phi_{w,n}$.

Proof. Since $\phi_{w,n}$ is already a representation of $W_{\mathbf{R}}$, the first induction is isomorphic to $\phi_{w,n} \oplus \phi_{w,n} \otimes \mathrm{sgn}$. Yet $\phi_{w,n}$ is itself induced from $\mathbf{C}^\times$, and so $\phi_{w,n} \otimes \mathrm{sgn} \simeq \phi_{w,n}$. The second claim was already noted above. Now suppose that $\varphi : W_{\mathbf{C}} \to \mathrm{GL}_2(\mathbf{C})$ is a complex L-parameter. All such parameters are of the form

$$z^{a_1}|z|^{-a_1}|z|^{w_1} \oplus z^{a_2}|z|^{-a_2}|z|^{w_2}$$

for integers a_1 and a_2. The induction of this representation to $W_{\mathbf{R}}$ is $\phi_{w_1,a_1} \oplus \phi_{w_2,a_2}$. Now consider the equality of $\mathrm{GL}_4(\mathbf{C})$-representations

$$\phi_{w_1,a_1} \oplus \phi_{w_2,a_2} = \phi_{(w;n,0)} = \phi_{w,n} \oplus \phi_{w,n}.$$

Restricting to $S^1 \subset \mathbf{C}^\times \subset W_{\mathbf{R}}$, we deduce that $|a_1| = |a_2| = n$, and then restricting to the action of $\mathbf{C}^\times$ on the eigenspace where $S^1 \subset \mathbf{C}^\times$ acts by z^n (which is distinct from z^{-n}), we deduce that $w_1 = w_2 = w$, and thus $\phi_{w_1,a_1} = \phi_{w_2,a_2} = \phi_{w,n}$. If a_1 and a_2 have opposite signs, then $\varphi = \phi_{w,n}$; otherwise we get the possibilities outlined in the statement of the lemma. Finally, (4) is immediate from the irreducibility of $\phi_{w,n}$. □

We note in passing that the $\mathrm{GSp}_4(\mathbf{C})$-parameter cannot be recovered, in general, from the $\mathrm{GL}_4(\mathbf{C})$-parameter. This is true in particular for $\phi_{(0;1,0)}$, since one may compute that the $\mathrm{GL}_4(\mathbf{C})$ representation preserves two symplectic forms whose similitude characters differ by sgn.

If K is a number field and $\boldsymbol{\pi}$ is an automorphic representation of $\mathrm{GL}_2(\mathbf{A}_K)$, we say that $\boldsymbol{\pi}$ has weight 0 if for each place $v|\infty$ of K, $\boldsymbol{\pi}_v$ corresponds to $\phi_{0,1}$. If F is a totally real field and π is an automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$, then we say that π has weight $(k_v, l_v)_{v|\infty}$ if for each place $v|\infty$ of F, we have $k_v \geq l_v \geq 2$ and $k_v \equiv l_v \pmod 2$, and π_v is in the L-packet corresponding to $\phi_{(2;k_v-1,l_v-2)}$. We say that π has parallel weight 2 if it has weight $(2,2)_{v|\infty}$ (we note that the congruence $k_v \equiv l_v \pmod 2$ is imposed in order to ensure that π is algebraic.)

2.7. **Galois representations associated to automorphic representations.** We now recall some results from [Mok14] on the existence of Galois representations (adapted to the particular setting of interest for us), beginning with the existence of Galois representations for certain cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$. The following theorem is essentially due to Sorensen [Sor10], although at the time that [Sor10] was written, some additional assumptions needed to be made, due to the lack of unconditional results on the transfer of automorphic representations between GSp_4 and GL_4.

Theorem 2.7.1. *Suppose that* F *is a totally real field, and that* π *is a cuspidal automorphic representation of* $\mathrm{GSp}_4(\mathbf{A}_F)$ *of weight* $(k_v, l_v)_{v|\infty}$, *where* $k_v \geq l_v > 2$ *and* $k_v \equiv l_v \pmod 2$ *for all* $v|\infty$. *Suppose also that* π *has central character* $|\cdot|^2$.

Fix a prime p. Then there is a continuous semisimple representation $\rho_{\pi,p} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ satisfying the following properties.

(1) $\nu \circ \rho_{\pi,p} = \varepsilon^{-1}$.

(2) For each finite place v, we have

$$\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}.$$

If furthermore $\rho_{\pi,p}$ is irreducible, then

$$\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2}).$$

(3) If $v|p$, then $\rho_{\pi,p}|_{G_{F_v}}$ is de Rham with Hodge–Tate weights $((k_v + l_v)/2 - 1, (k_v - l_v)/2 + 1, -(k_v - l_v)/2, 2 - (k_v + l_v)/2)$.

(4) If $\rho_{\pi,p}$ is irreducible, then for each finite place v of F, $\rho_{\pi,p}|_{G_{F_v}}$ is pure.

Proof. The existence of a representation $\rho_{\pi,p}$ valued in $\mathrm{GL}_4(\overline{\mathbf{Q}}_p)$ and satisfying (2) and (3) is part of [Mok14, Thm. 3.5] (note that the results of [Art04] cited in [Mok14] hold unconditionally by [GT19]). That the representation actually takes values in $\mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ with the claimed multiplier follows from [BC11, Cor. 1.3] (cf. [Mok14, Rem. 3.3(3)]). Finally, for part (4), note that if $\rho_{\pi,p}$ is irreducible, then π is of general type in the sense of [Art04] (see Section 2.9), and thus corresponds to an essentially self-dual algebraic automorphic representation Π of GL_4. Purity then follows from the main results of [Car12, Car14]. □

For representations which are ordinary in the sense of Section 2.4.23, we have the following variant on Theorem 2.7.1.

Theorem 2.7.2. *Suppose that F is a totally real field, and that π is a cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$ of weight $(k_v, l_v)_{v|\infty}$, where $k_v \geq l_v > 2$ and $k_v \equiv l_v \pmod 2$ for all $v|\infty$. Suppose also that π has central character $|\cdot|^2$.*

Fix a prime p. Assume that π_v is unramified at all places $v|p$, and that π is ordinary, with ordinary Hecke parameters $(\alpha_v, \beta_v)_{v|p}$. Then there is a continuous semisimple representation $\rho_{\pi,p} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ satisfying the following properties.

(1) $\nu \circ \rho_{\pi,p} = \varepsilon^{-1}$.

(2) For each finite place $v \nmid p$, we have

$$\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}.$$

If furthermore $\rho_{\pi,p}$ is irreducible, then

$$\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2}).$$

(3) If $v|p$, then

$$\rho_{\pi,p}|_{G_{F_v}} \cong \begin{pmatrix} \lambda_{\alpha_v}\varepsilon^{(k_v+l_v)/2-2} & * & * & * \\ 0 & \lambda_{\beta_v}\varepsilon^{(k_v-l_v)/2} & * & * \\ 0 & 0 & \lambda_{\beta_v}^{-1}\varepsilon^{-1-(k_v-l_v)/2} & * \\ 0 & 0 & 0 & \lambda_{\alpha_v}^{-1}\varepsilon^{1-(k_v+l_v)/2} \end{pmatrix}.$$

(4) If $\rho_{\pi,p}$ is irreducible, then for each finite place v of F, $\rho_{\pi,p}|_{G_{F_v}}$ is pure.

Proof. This follows from Theorem 2.7.1; part (3) is a standard consequence of p-adic Hodge theory, and is in particular immediate from [Ger19, Lem. 2.32] (and Proposition 2.4.6). □

The following theorem is a variant of the main result of [Mok14], which proves the existence of Galois representations associated to certain automorphic representations of $\mathrm{GL}_2(K)$, K a CM field.

Theorem 2.7.3. *Let F be a totally real field, and let K/F be a quadratic extension. Write $\mathrm{Gal}(K/F) = \{1, \tau\}$. Suppose that $\boldsymbol{\pi}$ is a cuspidal automorphic representation of $\mathrm{GL}_2(K)$ of weight 0 with trivial central character.*

Then there is a continuous irreducible representation $\rho_{\boldsymbol{\pi},p} : G_K \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ such that for each finite place $w \nmid p$ of K, we have

$$\mathrm{WD}(\rho_{\boldsymbol{\pi},p}|_{G_{K_w}})^{\mathrm{ss}} \cong \mathrm{rec}_p(\boldsymbol{\pi}_w \otimes |\cdot|^{-1/2})^{\mathrm{ss}}.$$

If $\boldsymbol{\pi}_w$ is not a twist of a Steinberg representation, then in fact

$$\mathrm{WD}(\rho_{\boldsymbol{\pi},p}|_{G_{K_w}})^{F-\mathrm{ss}} \cong \mathrm{rec}_p(\boldsymbol{\pi}_w \otimes |\cdot|^{-1/2}).$$

For each place $w|p$ of K, the representation $\rho_{\boldsymbol{\pi},p}|_{G_{K_v}}$ is Hodge–Tate, and for each $\tau : K \hookrightarrow \overline{\mathbf{Q}}_p$, the τ-labelled Hodge–Tate weights of $\rho_{\boldsymbol{\pi},p}$ are $(0, 1)$.

Proof of Theorem 2.7.3. In the case that K is CM this is a special case of the main theorem of [Mok14], and essentially the same proof works in the general case. The argument of [Mok14, §5.1] goes over unchanged to produce a cuspidal automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ (see Theorem 2.9.3 below); to see that π_v is in the L-packet corresponding to $\phi_{(2;1,0)}$ at each place $v|\infty$ of F, one uses Lemma 2.6.1 at the places which split in K, and [Mok14, Prop. 5.2] at the places for which K_v is complex. One then easily checks that the arguments of [Mok14, §5.2-5.3] go over without any changes to the case of general K, as required. □

2.8. **Compatible systems of Galois representations, L-functions, and Hasse–Weil zeta functions.** We now recall some definitions concerning compatible systems from [BLGGT14b, §5] and [PT15, §1]; in fact, our definition of a "strictly compatible system" differs slightly from the definitions in those papers, because we find it convenient to include local-global compatibility at places dividing p. Let F denote a number field. By a *rank n weakly compatible system of l-adic representations $\mathcal{R}$ of G_F defined over M* we mean a 5-tuple

$$(M, S, \{Q_v(X)\}, \{r_\lambda\}, \{H_\tau\})$$

where

(1) M is a number field considered as a subfield of $\mathbf{C}$;
(2) S is a finite set of primes of F;
(3) for each prime $v \notin S$ of F, $Q_v(X)$ is a monic degree n polynomial in $M[X]$;
(4) for each prime λ of M (with residue characteristic l, say)

$$r_\lambda : G_F \longrightarrow \mathrm{GL}_n(\overline{M}_\lambda)$$

is a continuous, semi-simple, representation such that
- if $v \notin S$ and $v \nmid l$ is a prime of F, then r_λ is unramified at v and $r_\lambda(\mathrm{Frob}_v)$ has characteristic polynomial $Q_v(X)$,
- while if $v|l$, then $r_\lambda|_{G_{F_v}}$ is de Rham and in the case $v \notin S$ crystalline;

(5) for $\tau : F \hookrightarrow \overline{M}$, H_τ is a multiset of n integers such that for any $\overline{M} \hookrightarrow \overline{M}_\lambda$ over M we have $\mathrm{HT}_\tau(r_\lambda) = H_\tau$.

If $\mathcal{R} = (M, S, \{Q_v(X)\}, \{r_\lambda\}, \{H_\tau\})$ and $\mathcal{R}' = (M', S', \{Q'_v(X)\}, \{r'_\lambda\}, \{H'_\tau\})$ are two compatible systems, then we write $\mathcal{R} \cong \mathcal{R}'$ if $Q_v(X) = Q'_v(X)$ for a set of places v of Dirichlet density one. This implies that $Q_v(X) = Q'_v(X)$ for all $v \notin S \cup S'$, and that $r_\lambda \cong r'_\lambda$ for all λ, and $H'_\tau = H_\tau$ for all τ.

We say that $\mathcal{R}$ is *regular* if for each $\tau : F \hookrightarrow \overline{M}$, the elements of H_τ are pairwise distinct. We will call $\mathcal{R}$ *strictly compatible* if for each finite place v of F there is a Weil–Deligne representation $\mathrm{WD}_v(\mathcal{R})$ of W_{F_v} over $\overline{M}$ such that for each place λ of M and every M-linear embedding $\varsigma : \overline{M} \hookrightarrow \overline{M}_\lambda$ we have $\varsigma\,\mathrm{WD}_v(\mathcal{R}) \cong \mathrm{WD}(r_\lambda|_{G_{F_v}})^{\text{F-ss}}$.

We will call a strictly compatible system $\mathcal{R}$ *pure* of weight w if for each finite place v of F the Weil–Deligne representation $\mathrm{WD}_v(\mathcal{R})$ is pure of weight w.

The following result is well-known (see for example [Fon94, Rem. 2.4.6]), but as we do not know of a convenient reference for a proof, we briefly explain how it follows from results in the literature.

Proposition 2.8.1. *If A is an abelian variety over a number field F, then, for each $0 \leq i \leq 2\dim X$, the l-adic cohomology groups $H^i(A_{\overline{F}}, \overline{\mathbf{Q}}_l)$ form a strictly compatible system which is pure of weight i and which is defined over $\mathbf{Q}$.*

Proof. Since $H^i(A, \overline{\mathbf{Q}}_l) = \wedge^i H^1(A, \overline{\mathbf{Q}}_l)$, it is enough to check the case $i = 1$. The compatible system satisfies strict compatibility at the places not dividing l by [Noo13, Cor. 2.7]. In the case that A has semistable reduction, it is furthermore strictly compatible by [Noo17, Cor. 2.2]. One can deduce the general case from this by a base change trick due to Saito [Sai97], which was exploited in [Kis08, Ski09, BLGGT14a]. Indeed, as in the proof of [BLGGT14a, Thm. 2.1], it suffices for each finite place v of F to check that whenever $g \in W_{F_v}$ maps to a positive power of Frobenius in the absolute Galois group of the residue field, then the trace of g on $\mathrm{WD}(H^1(A, \overline{\mathbf{Q}}_l))$ is independent of l. One can choose an extension E/F (for example, the fixed field of the subgroup of W_F generated by g and the kernel of the restriction to I_F of $\mathrm{WD}(H^1(A, \overline{\mathbf{Q}}_l))$ for some l) and a place $v|w$ of E such that A_E is semistable and $g \in W_{E_w}$, and the claim then follows from the independence of l for A_E.

It remains to check purity. By [Ray94, Thm. 4.2.2], it is enough to check purity for the Weil–Deligne representations associated to 1-motives with potentially good reduction, which is [Ray94, Prop. 4.6.1, Prop. 4.7.4].

The above is of course not a historically accurate account of a proof; indeed, the strict compatibility of the compatible system at places not dividing l is stated in [Del73, Ex. 8.10], and given Fontaine's definition of the Weil–Deligne representation associated to a potentially semistable representation, the entire proposition can be deduced from the results of [GRR72]. We omit the details, but we would like to thank Brian Conrad for explaining them to us. □

Definition 2.8.2. If A/F is an abelian surface, then we write $\rho_{A,l}$ for $H^1(A_{\overline{F}}, \overline{\mathbf{Q}}_l)$, and $\mathcal{R}_A$ for the compatible system $\{\rho_{A,l}\}$. We can think of $\rho_{A,l}$ as a representation $\rho_{A,l} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_l)$ with multiplier ε_l^{-1}, and will frequently do so without comment.

Remark 2.8.3. It will sometimes be convenient to say that a set of GSp_4-valued representations form a compatible system, by which we simply mean that the corresponding GL_4-valued representations form a compatible system. In particular, the representations $\rho_{A,l} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_l)$ considered in Definition 2.8.2 form a

compatible system in this sense. (In general, one might wish to ask for a compatibility between the symplectic structures; such a compatibility always holds in the cases that we consider, and in particular we will only consider representations whose multiplier character is the inverse cyclotomic character, so we ignore this point.)

We can define the L-function of $\mathcal{R}$ as follows:

$$L(\mathcal{R}, s) = \prod_{v \nmid \infty} L(\mathrm{WD}_v(\mathcal{R}), s).$$

Furthermore, if $\mathcal{R}$ comes from an abelian variety (or more generally, arises in a geometric structure where the Hodge structure is apparent) then (as in [Ser70]) we can define Gamma factors $L_v(\mathcal{R}, s)$ for each place $v|\infty$ of F, and we set

$$\Lambda(\mathcal{R}, s) = L(\mathcal{R}, s) \prod_{v|\infty} L_v(\mathcal{R}, s). \tag{2.8.4}$$

In particular, if $\mathcal{R}$ arises from an abelian surface over a totally real field F, then the corresponding Gamma factor is given by $L_v(\mathcal{R}, s) = \Gamma_{\mathbf{C}}(s)^2$ for all $v|\infty$ where $\Gamma_{\mathbf{C}}(s) = (2\pi)^{-s}\Gamma(s)$.

We also have a conductor $N(\mathcal{R})$ which is a product of local factors depending only on the $\mathrm{WD}_v(\mathcal{R})$. Conjecturally, if $\mathcal{R}$ is a strictly compatible system, then $\Lambda(\mathcal{R}, s)$ admits a meromorphic continuation to the entire complex plane and satisfies a functional equation of the form

$$\Lambda(\mathcal{R}, s) = \varepsilon(\mathcal{R}) N(\mathcal{R})^{-s} \Lambda(\mathcal{R}^\vee, 1 - s) \tag{2.8.5}$$

for some factor $\varepsilon(\mathcal{R})$. (When $\mathcal{R}$ arises geometrically, there are natural definitions of the epsilon factor $\varepsilon(\mathcal{R})$, but it is not immediately apparent how to read off $\varepsilon(\mathcal{R})$ directly from the compatible system.)

In particular, if A/F is an abelian variety, then by Proposition 2.8.1

$$\Lambda_i(A, s) := \Lambda(H^i(A_{\overline{F}}, \overline{\mathbf{Q}}_l), s)$$

is well-defined, and we define the completed Hasse–Weil zeta function of A to be

$$\Lambda(A, s) := \prod_{i=0}^{2\dim A} \Lambda_i(A, s)^{(-1)^i}.$$

Note that if v is a finite place of F at which A has good reduction with corresponding reduction $\overline{A}$, then the local L-factor

$$L_v(A, s) := \prod_{i=0}^{2\dim A} L(\mathrm{WD}(H^i(A_{\overline{F}}, \overline{\mathbf{Q}}_l)), s)^{(-1)^s}$$

can be written as

$$L_v(A, s) = \exp\left(\sum_{m=1}^{\infty} \frac{\#\overline{A}(k(v)_m)}{m} \#k(v)^{-ms}\right)$$

where $k(v)$ is the residue field of F_v and $k(v)_m/k(v)$ is the extension of degree m.

We have the following conjectures for the $\Lambda_i(A, s)$, which we will prove for abelian surfaces over totally real fields by showing that they are potentially automorphic.

Conjecture 2.8.6 ([Ser70], Conj. C9)**.** *For each i, $\Lambda_i(A,s)$ has a meromorphic continuation to the entire complex plane, and satisfies a functional equation of the form*

$$\Lambda_i(A,s) = wN^{\frac{i+1}{2}-s}\Lambda_i(A, i+1-s)$$

where $w = \pm 1$ and $N \in \mathbf{Z}_{\geq 1}$.

Corollary 2.8.7. *If Conjecture 2.8.6 holds, then $\Lambda(A,s)$ has a meromorphic continuation to the entire complex plane, and satisfies a functional equation of the form $\Lambda(A,s) = \varepsilon N^{-s}\Lambda(A, 1+\dim A - s)$ where $\varepsilon \in \mathbf{R}$ and $N \in \mathbf{Q}_{>0}$.*

Proof. This follows immediately from Conjecture 2.8.6 by Poincaré duality. □

2.9. **Arthur's classification.** We now recall some consequences of Arthur's classification [Art04] of discrete automorphic representations of GSp_4. The analogous classifications for Sp_4 and SO_5 are special cases of the very general results proved in [Art13], and a proof of the classification announced in [Art04], making use of the results and techniques of [Art13] is given in [GT19]. This reference establishes the compatibility of Arthur's classification with the local Langlands correspondence $\mathrm{rec}_{\mathrm{GT}}$, which we use below without further comment.

We say that an automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ is discrete if it occurs in the discrete spectrum of the L^2-automorphic forms (with fixed central character $\omega = \omega_\pi$). Note in particular that all cuspidal automorphic representations are discrete. Arthur's classification divides the discrete spectrum into six families of automorphic representations. We will not need the full details of this classification, but rather just some consequences that we now recall.

If Π is a cuspidal automorphic representation of $\mathrm{GL}_4(\mathbf{A}_F)$, then we say that Π is of *symplectic type* with multiplier χ if the partial L-function $L^S(s, \Pi, \bigwedge^2 \otimes \chi^{-1})$ has a pole at $s = 1$ (where S is any finite set of places of F). Note that this implies in particular that $\Pi \cong \Pi^\vee \otimes \chi$.

We say that a discrete automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ is *of general type* in the sense of [Art04] if there is a cuspidal automorphic representation Π of $\mathrm{GL}_4(\mathbf{A}_F)$ of symplectic type with multiplier ω_π such that for each place v of F, the L-parameter obtained from $\mathrm{rec}_{\mathrm{GT}}(\pi_v)$ by composing with the usual embedding $\mathrm{GSp}_4 \hookrightarrow \mathrm{GL}_4$ is $\mathrm{rec}(\Pi_v)$. We say that Π is the *transfer* of π.

In practice, all of the automorphic representations π that we consider in our main arguments will be of general type. We will often use the following lemma to guarantee this. (For example, the lemma will be used to show that when we localize a cohomology group at a non-Eisenstein maximal ideal, the only automorphic representations that contribute are of general type.)

Lemma 2.9.1. *Suppose that F is totally real, and that π is a discrete automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$, and that at each place $v|\infty$, π_v has the same infinitesimal character as the representations in the L-packet corresponding to $\varphi_{(2;k_v-1,l_v-2)}$ with $k_v \equiv l_v \pmod 2$ and $k_v \geq l_v \geq 2$. Suppose that π is not of general type.*

Then there is a compatible system of reducible *Galois representations $\rho_{\pi,p} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ such that for all but finitely many places v of F, we have $\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}$.*

Proof. We follow the proof of [CG20, Thm. 7.11]. Since π is not of general type, π falls into one of the five classes (b)-(f) listed at the end of [Art04]. In cases (e) and (f), we see that the Hecke parameters of π agree with those of a direct sum of 4

idele class characters. By the hypothesis on the infinitesimal character, these characters are algebraic, so we may take the direct sum of the corresponding compatible systems of Galois representations.

In case (d), the Hecke parameters of π agree with those of an isobaric direct sum of the form $\lambda|\cdot|^{1/2} \boxplus \lambda|\cdot|^{-1/2} \boxplus \mu$, where λ is an idele class character, and μ is a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_F)$, satisfying $\omega_\mu = \lambda^2 = \omega_\pi$. Considering infinitesimal characters, we see that λ is algebraic, so that $\lambda|\cdot|^{1/2} \boxplus \lambda|\cdot|^{-1/2}$ is regular algebraic. This implies that μ is also regular algebraic, and thus has an attached compatible system of Galois representations.

In case (b), the Hecke parameters of π agree with those of an isobaric direct sum of the form $\mu_1 \boxplus \mu_2$, where $\mu_1 \neq \mu_2$ are cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_F)$ with central character μ_π. Since their central characters agree, it follows easily that they both correspond to holomorphic Hilbert modular eigenforms of paritious weight. Finally in case (c), the Hecke parameters of π agree with those of an isobaric direct sum of the form $\mu|\cdot|^{1/2} \boxplus \mu|\cdot|^{-1/2}$, where μ is a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_F)$ of orthogonal type; that is, it is induced from a quadratic extension of F. Since μ is certainly algebraic, we again have an attached compatible system of reducible Galois representations, as required. □

Remark 2.9.2. Suppose that π is of general type but otherwise satisfies the conditions of Lemma 2.9.1. Then the corresponding Galois representations constructed in [Mok14] (see also Theorem 2.7.1) give rise to a compatible system of Galois representations which — in contrast to those occurring in Lemma 2.9.1 — are expected to always be irreducible.

The following theorem summarizes the consequences that we need from Arthur's multiplicity formula.

Theorem 2.9.3. *Suppose that F is a totally real field, and that Π is a cuspidal automorphic representation of $\mathrm{GL}_4(\mathbf{A}_F)$ of symplectic type with multiplier χ. Then there exists at least one discrete automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ with central character χ such that Π is the transfer of π.*

More precisely, for each place v of F, let π_v be an element of the L-packet corresponding to $(\mathrm{rec}_p(\Pi_v), \chi_v)$. Then $\pi := \otimes'_v \pi_v$ is automorphic, and occurs with multiplicity one in the discrete spectrum.

If, furthermore, Π is algebraic, then π is cuspidal.

Proof. The statements of the first two paragraphs are immediate from the multiplicity formula of [Art04] as proved in [GT19] (note that since π is of general type by definition, the group $\mathcal{S}_\psi$ considered in [Art04] is trivial). Suppose then that Π is algebraic; then Π_∞ is essentially tempered by [Clo90, Lem. 4.9], so that π_∞ is also essentially tempered (as its L-parameter is essentially bounded), so that π is cuspidal by [Wal84, Thm. 4.3]. □

2.10. **Balanced modules.** Let S be a Noetherian local ring with residue field k, and let M be a finitely generated S-module. As in [CG18, §2.1], we define the *defect* $d_S(M)$ to be

$$d_S(M) := \dim_k M/\mathfrak{m}_S M - \dim_k \mathrm{Tor}^1_S(M, k).$$

Definition 2.10.1. We say that M is *balanced* if $d_S(M) \geq 0$.

Lemma 2.10.2. *If M is balanced, then there is a presentation*

$$S^d \to S^d \to M \to 0$$

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

Conversely if M admits a presentation

$$S^r \to S^r \to M \to 0$$

for some $r \geq 0$, then M is balanced.

Proof. Assume firstly that M is balanced, and choose a (possibly infinite) minimal resolution

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

by finite free S-modules P_i of rank r_i. (Recall that a minimal resolution is one whose differentials vanish modulo $\mathfrak{m}_S$, and that such a resolution always exists.) Tensoring this resolution with k over S, we see that $r_i = \dim_k \mathrm{Tor}_S^i(M, k)$, so that in particular by our assumptions we have $d = r_0 \geq r_1$, so that there is a presentation of the form $P_1 \oplus S^{\oplus(d-r_1)} \to P_0 \to M \to 0$, as required.

Conversely, if M admits a presentation $S^r \to S^r \to M \to 0$, then let K be the image of the map $S^r \to S^r$. Then from the exact sequence

$$0 \to \mathrm{Tor}_S^1(M, k) \to K/\mathfrak{m}_S K \to k^r \to M/\mathfrak{m}_S M \to 0$$

we see that

$$d_S(M) = r - \dim_k K/\mathfrak{m}_S K;$$

since K admits a surjection from S^r, it follows that $d_S(M) \geq 0$, as required. □

2.11. **Projectors.** Let R be a complete local Noetherian ring with maximal ideal $\mathfrak{m}_R$ and finite residue field. We let $\mathbf{Mod}^{\mathrm{comp}}(R)$ be the category of $\mathfrak{m}_R$-adically complete and separated R-modules. Let $M \in \mathrm{Ob}(\mathbf{Mod}^{\mathrm{comp}}(R))$ and $T \in \mathrm{End}_R(M)$.

Definition 2.11.1. We say that T is locally finite on M if for all $n \geq 0$, $M/\mathfrak{m}_R^n$ is an inductive limit of finite type R-modules which are stable under the action of T.

Lemma 2.11.2. *If T_1, T_2 commute and are both locally finite on M, then T_1T_2 is also locally finite on M.*

Proof. By definition we can assume that M is $\mathfrak{m}_R^n$-torsion for some n. If $v \in M$ then since T_1 is locally finite, the R-submodule of M generated by the $T_1^i v$ is finitely generated. Since T_2 is locally finite, it follows that the R-submodule generated by the $T_2^i T_1^j v$ is also finitely generated, and since T_1, T_2 commute, this submodule is stable under the action of T_1T_2, as required. □

The following results from [Pil20] will be used to construct the ordinary projectors associated to certain Hecke operators.

Lemma 2.11.3 ([Pil20, Lem. 2.1.2])**.** *If M is an object of $\mathbf{Mod}^{\mathrm{comp}}(R)$ and T is an endomorphism of M, then T is locally finite on M if and only if it is locally finite on $M/\mathfrak{m}_R$.*

Lemma 2.11.4 ([Pil20, Lem. 2.1.3])**.** *If T is locally finite on M, then $\lim_{n\to\infty} T^{n!}$ converges pointwise in the $\mathfrak{m}_R$-adic topology to a projector $e(T)$ on M.*

The operators T and $e(T)$ commute, and we have a T-stable decomposition

$$M = e(T)M \oplus (1 - e(T))M,$$

where T is bijective on $e(T)M$ and topologically nilpotent on $(1 - e(T))M$.

We call $e(T)$ the *ordinary projector* attached to T. Let $\mathbf{D}(R)$ be the derived category of R-modules, let $\mathbf{D}^{\mathrm{flat}}(R)$ be the full subcategory of $\mathbf{D}(R)$ generated by bounded complexes of flat, $\mathfrak{m}_R$-adically complete and separated R-modules and let $\mathbf{D}^{\mathrm{perf}}(R)$ be the full subcategory of $\mathbf{D}(R)$ generated by bounded complexes of finite free R-modules. Let $M \in \mathrm{Ob}(\mathbf{D}^{\mathrm{flat}}(R))$. We say that an operator $T \in \mathrm{End}(M)$ is locally finite if there is a bounded complex of flat modules N representing M and an operator $T_0 \in \mathrm{End}(N)$ representing T which is degree-wise locally finite. By [Pil20, Lem. 2.3.1], T is locally finite on M if and only if T is locally finite on the cohomology groups $H^i(M \otimes_R^L R/\mathfrak{m}_R)$ and there is a bounded complex of flat modules N representing M and an operator $T_0 \in \mathrm{End}(N)$ representing T. Given a choice of representatives $(N, T_0 \in \mathrm{End}(N))$ for a locally finite operator T, we get an associated idempotent $e(T_0) \in \mathrm{End}(N)$. In general, we do not know whether two choices of representatives $(N, T_0 \in \mathrm{End}(N))$ give the same projector in $\mathrm{End}_{\mathbf{D}(R)}(M)$. But by [Pil20, Lem. 2.3.2], if we assume that for one choice of representative $e(T_0)M$ is an object of $\mathbf{D}^{perf}(R)$ then, for another choice of locally finite representative $(N', T_1 \in \mathrm{End}(N'))$, $e(T_1)M$ is an object of $\mathbf{D}^{perf}(R)$ and there is a canonical quasi-isomorphism $e(T_0)M \to e(T_1)M$. In the sequel, these conditions will always be satisfied and we will write $e(T)$ by abuse of notation.

3. Shimura varieties

In this section, we discuss the Hilbert–Siegel Shimura varieties that we work with, and some properties of their integral models. There are two closely related algebraic groups here: $G_1 = \mathrm{Res}_{F/\mathbf{Q}}\mathrm{GSp}_4$ and its subgroup G of elements with similitude factor in $\mathbf{G}_m \hookrightarrow \mathrm{Res}_{F/\mathbf{Q}}\mathbf{G}_m$.

The group G admits a standard PEL Shimura variety and there is a good moduli interpretation, integral models, and a good theory of integral compactification. Nonetheless, from an automorphic view point we must work with the group G_1 which gives rise to a Shimura variety of abelian type.

Going back to the work of Deligne (see in particular [Del79, §2.7]), there is a standard strategy for handling abelian type Shimura varieties by relating their connected components to quotients of connected components of Hodge type Shimura varieties by finite groups. As a particular instance of this strategy, the Shimura varieties for G and G_1 are closely related: the connected components of G_1-Shimura varieties are quotients of the connected components of G-Shimura varieties by finite groups. We therefore study both of them at the same time.

For convenience, our main references for integral models of PEL Shimura varieties and their compactifications are the papers [Lan13, Lan16, Lan17], although some of the results we cite from there were proved in earlier papers, in particular [Kot92]; we refer the reader to the references in [Lan13] for a more detailed historical account.

3.1. **Similitude groups.** Let F be a totally real field. Let $V = \mathcal{O}_F^4$ be a free $\mathcal{O}_F$-module of rank 4. We equip V with the symplectic $\mathcal{O}_F$-linear form $<,>_1 \colon V \times V \to \mathcal{O}_F$ given by the matrix J. We let $<,> := (\mathrm{Tr}_{F/\mathbf{Q}} \circ <,>_1)$ be the associated $\mathbf{Z}$-linear symplectic form.

Let $G_1 = \mathrm{Res}_{F/\mathbf{Q}}\mathrm{GSp}_4$ be the algebraic group of symplectic F-linear automorphisms of $(V_{\mathbf{Q}}, <,>_1)$, up to a similitude factor ν in $\mathrm{Res}_{F/\mathbf{Q}}\mathbf{G}_m$.

Let $G \subset G_1$ be the algebraic group of symplectic F-linear automorphisms of $(V_{\mathbf{Q}}, <,>)$ up to a similitude factor in $\mathbf{G}_m$; that is, $G = G_1 \times_{\nu, \mathrm{Res}_{F/\mathbf{Q}}\mathbf{G}_m} \mathbf{G}_m$.

3.2. **Shimura varieties over C.** We firstly briefly discuss some Shimura varieties over $\mathbf{C}$. We caution the reader that in the bulk of the paper we will work with Shimura varieties over $\mathbf{Z}_{(p)}$ which are not quite integral models of these Shimura varieties, but whose geometrically connected components are the same as these; see Proposition 3.3.9 below for a precise statement. We begin by recalling the definition of a neat compact open subgroup from [Lan13, Defn. 1.4.1.8].

Definition 3.2.1. Write $g = (g_l)_l \in G_1(\mathbf{A}^\infty)$, and for each l, write Γ_{g_l} for the subgroup of $\overline{\mathbf{Q}}_l^\times$ generated by the eigenvalues of g_l (under any faithful linear representation of G_1). Then we say that g is *neat* if

$$\bigcap_l (\overline{\mathbf{Q}}^\times \cap \Gamma_{g_l})_{\mathrm{tors}} = 1.$$

Similarly, if $g \in G_1(\mathbf{A}^{\infty,p})$, then we say that g is neat if

$$\bigcap_{l \neq p} (\overline{\mathbf{Q}}^\times \cap \Gamma_{g_l})_{\mathrm{tors}} = 1.$$

We say that a compact open subgroup $K \subset G_1(\mathbf{A}^\infty)$ (resp. $K^p \subset G_1(\mathbf{A}^{\infty,p})$) is *neat* if all of its elements are neat.

We consider the Shimura variety associated to the group G_1 and a neat compact open subgroup $K \subset G_1(\mathbf{A}^\infty)$:

$$S_K^{G_1}(\mathbf{C}) = G_1(\mathbf{Q}) \backslash \big(G_1(\mathbf{R}) \times G_1(\mathbf{A}^\infty)\big) / Z(\mathbf{R})^0 K_\infty^0 K$$

where $Z(\mathbf{R})^0 \simeq \mathbf{R}_{>0}^{\mathrm{Hom}(F,\mathbf{R})}$ is the connected component of the centre in $G_1(\mathbf{R}) \simeq \mathrm{GSp}_4(\mathbf{R})^{\mathrm{Hom}(F,\mathbf{R})}$ and K_∞^0 is the connected component of the maximal compact subgroup inside $G_1(\mathbf{R})$, so that K_∞^0 is a product of copies of $U(2)$. This Shimura variety carries a natural structure of complex quasi-projective variety, as we have $G_1(\mathbf{R})/Z(\mathbf{R})^0 K_\infty^0 = (\mathcal{H} \cup -\mathcal{H})^{\mathrm{Hom}(F,\mathbf{R})}$, where $\mathcal{H}$ is the Siegel half space of symmetric matrices $M = A + iB \in \mathrm{M}_{2\times 2}(\mathbf{C})$ with B positive definite.

Let $G_1(\mathbf{Q})^+$ be the subgroup of $G_1(\mathbf{Q})$ equal to $\nu^{-1}(F^{\times,+})$, where $F^{\times,+}$ is the subgroup of totally positive elements in $F^\times$. Then by strong approximation,

$$G_1(\mathbf{A}^\infty) = \coprod_c G_1(\mathbf{Q})^+ cK$$

where c runs through a (finite) set of elements in $G_1(\mathbf{A}^\infty)$ such that $\nu(c)$ are representatives of the strict class group $F^{\times,+} \backslash (\mathbf{A}^\infty \otimes_{\mathbf{Q}} F)^\times / \nu(K)$.

One can then write

$$S_K^{G_1}(\mathbf{C}) = \coprod_c \Gamma_1(c, K) \backslash \mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$$

where $\Gamma_1(c, K) = G_1(\mathbf{Q})^+ \cap cKc^{-1}$.

This Shimura variety, although natural from the point of view of automorphic forms, is not of PEL type. Therefore, it is also necessary to work with another Shimura variety. We can consider the double quotient

$$S_K^G(\mathbf{C}) = G(\mathbf{Q}) \backslash \big(G(\mathbf{R}) \times G_1(\mathbf{A}^\infty)\big) / \mathbf{R}_{>0} K_\infty^0 K;$$

this is not strictly speaking a Shimura variety, and in particular we emphasise that it is not the PEL Shimura variety associated to G! By strong approximation we

may write

$$G_1(\mathbf{A}^\infty) = \coprod_c G(\mathbf{Q})^+ cK$$

where c runs through a set of elements of $G_1(\mathbf{A}^\infty)$ such that $\nu(c)$ are representatives of the infinite set $\mathbf{Q}^{\times,+}\backslash(\mathbf{A}^\infty \otimes_{\mathbf{Q}} F)^\times/\nu(K)$. For all c, we consider the group $\Gamma(c,K) = G(\mathbf{Q})^+ \cap cKc^{-1}$, so that

$$S_K^G(\mathbf{C}) = \coprod_c \Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}.$$

The inclusion $G(\mathbf{R}) \hookrightarrow G_1(\mathbf{R})$ induces a natural surjective map $S_K^G(\mathbf{C}) \to S_K^{G_1}(\mathbf{C})$. On connected components, it induces the natural map

$$\mathbf{Q}^{\times,+}\backslash(\mathbf{A}^\infty \otimes_{\mathbf{Q}} F)^\times/\nu(K) \to F^{\times,+}\backslash(\mathbf{A}^\infty \otimes_{\mathbf{Q}} F)^\times/\nu(K).$$

For any $c \in G_1(\mathbf{A}^\infty)$ we have an associated surjective map on the connected components corresponding to c, given by

$$\Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})} \to \Gamma_1(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}.$$

Let $Z(\Gamma_1(c,K)) \subset \Gamma_1(c,K)$ be the centre. Then $Z(\Gamma_1(c,K))$ is a finite index subgroup of $\mathcal{O}_F^\times$ that we denote by $\mathcal{O}_F^\times(K)$. Let

$$\Delta(K) = \Gamma_1(c,K)/\big(\mathcal{O}_F^\times(K), \Gamma(c,K)\big).$$

This is a finite group, independent of c and isomorphic to

$$\nu(\Gamma_1(c,K))/\nu(\mathcal{O}_F^\times(K)) = (\mathcal{O}_F^{\times,+} \cap \nu(K))/\nu(\mathcal{O}_F^\times(K)),$$

having noted the following:

Lemma 3.2.2. *There is an equality* $\nu(\Gamma_1(c,K)) = \mathcal{O}_F^{\times,+} \cap \nu(K)$.

Proof. Recall that by definition $\Gamma_1(c,K) = G_1(\mathbf{Q})^+ \cap cKc^{-1}$, so certainly we have an inclusion $\nu(\Gamma_1(c,K)) \subseteq \mathcal{O}_F^{\times,+} \cap \nu(K)$. Conversely, suppose that $\nu(\gamma) = x \in \mathcal{O}_F^{\times,+} \cap \nu(K)$ for some element $\gamma \in cKc^{-1}$. Since $x \in \mathcal{O}_F^{\times,+} = \nu(G_1(\mathbf{Q})^+)$, we can choose $g \in G_1(\mathbf{Q})^+$ with $\nu(g) = x$. Then $\nu(\gamma^{-1}g) = 1$, so by strong approximation, we may find an $h \in G_1(\mathbf{Q})^+$ with trivial similitude character such that h is arbitrarily close to $g\gamma^{-1}$, and in particular close enough that $h\gamma g^{-1}$ lies in cKc^{-1}. Then $h\gamma \in G_1(\mathbf{Q})^+ \cap cKc^{-1}$ and has similitude character x, as required. □

We also have

Lemma 3.2.3. *The map* $\Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})} \to \Gamma_1(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$ *is finite étale with group* $\Delta(K)$.

Proof. The group $\Gamma_1(c,K)$ acts through its quotient $\Gamma_1(c,K)/\mathcal{O}_F^\times(K)$ on $\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$, and since K is neat, this action is free. □

3.3. **Integral models of Shimura varieties.** We now introduce the integral models of Shimura varieties that we will consider in the rest of the paper.

3.3.1. *Compact open subgroups at p.* We let p be a prime that is totally split in F. Let v be a prime ideal in $\mathcal{O}_F$ above p. Consider the following chain of $\mathcal{O}_{F_v}$-sub modules of F_v^4:

$$V_0 \to V_1 \to V_2 \to V_3 \to V_4$$

where $V_0 = V \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v} = \oplus_{i=1}^4 \mathcal{O}_{F_v} e_i$ and $V_j = \langle p^{-1}e_1, \cdots, p^{-1}e_j, e_{j+1}, \cdots, e_4\rangle$. We can identify V_0 and V_4 through multiplication by p and sometimes think of the indices as being in $\mathbf{Z}/4\mathbf{Z}$.

From the perfect pairing $<,>$ on V_0 we obtain perfect pairings on $V_2 \times V_2$ and on $V_1 \times V_3$.

We now recall the definitions of the parabolic subgroups that we use in terms of flags; this description is well suited to the definitions of our integral models.

- $\mathrm{GSp}_4(\mathcal{O}_{F_v}) = \mathrm{Aut}(V_0) \cap \mathrm{GSp}_4(F_v)$ (the hyperspecial subgroup),
- $\mathrm{Par}(v) = \mathrm{Aut}(V_1 \to V_3) \cap \mathrm{GSp}_4(F_v)$ (the paramodular subgroup),
- $\mathrm{Si}(v) = \mathrm{Aut}(V_0 \to V_2) \cap \mathrm{GSp}_4(F_v)$ (the Siegel parahoric),
- $\mathrm{Kli}(v) = \mathrm{Aut}(V_0 \to V_1 \to V_3 \to V_0) \cap \mathrm{GSp}_4(F_v)$ (the Klingen parahoric),
- $\mathrm{Iw}(v) = \mathrm{Aut}(V_0 \to V_1 \to V_2 \to V_3 \to V_0) \cap \mathrm{GSp}_4(F_v)$ (the Iwahori subgroup).

3.3.2. *The moduli problem.* Let $ALG/\mathbf{Z}_{(p)}$ be the category of Noetherian $\mathbf{Z}_{(p)}$-algebras and $AFF/\mathbf{Z}_{(p)}$ the opposite category. Let $K \subset G_1(\mathbf{A}^\infty)$ be a compact open subgroup; we will also refer to such a compact open subgroup as a *level structure*.

Definition 3.3.3. We say that a level structure $K = K^pK_p$ is *reasonable* if $K^p \subset G(\mathbf{A}^{\infty,p})$ is neat, and if $K_p = \prod_{v|p} K_v$ where for each $v|p$ we have

$$K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v), \mathrm{Si}(v), \mathrm{Kli}(v), \mathrm{Iw}(v)\}.$$

Let K be a reasonable level structure. We consider the groupoid $\mathbf{Y}_K$ over $AFF/\mathbf{Z}_{(p)}$ whose fibre over $S = \operatorname{Spec} R \in \mathrm{Ob}(AFF/\mathbf{Z}_{(p)})$ is the category with objects $(A, \iota, \lambda, \eta, \eta_p)$, where:

(1) $A \to \operatorname{Spec} R$ is an abelian scheme,
(2) $\iota : \mathcal{O}_F \to \mathrm{End}(G) \otimes \mathbf{Z}_{(p)}$ is an action,
(3) $\mathrm{Lie}(A)$ is a locally free $\mathcal{O}_F \otimes_{\mathbf{Z}} R$-module of rank 2,
(4) $\lambda : A \to A^t$ is a prime to p, $\mathcal{O}_F$-linear quasi-polarization such that for all $v|p$, $\mathrm{Ker}(\lambda : A[v^\infty] \to A^t[v^\infty])$ is trivial if $K_v \neq \mathrm{Par}(v)$ and is an order p^2 group scheme if $K_v = \mathrm{Par}(v)$,
(5) η is a K^p-level structure,
(6) η_p is a K_p-level structure.

Here by a prime to p quasi-polarization $\lambda : A \to A^t$ we mean a $\mathbf{Z}_{(p)}^\times$-polarization in the sense of [Lan13, Defn. 1.3.2.19]. By a K_p-level structure η_p, we mean the following list of data:

(1) For all $v|p$ such that $K_v = \mathrm{Kli}(v)$, $H_v \subset A[v]$ is an order p-group scheme,
(2) For all $v|p$ such that $K_v = \mathrm{Si}(v)$, $L_v \subset A[v]$ is an order p^2 group scheme that is totally isotropic for the Weil pairing.
(3) For all $v|p$ such that $K_v = \mathrm{Iw}(v)$, $H_v \subset L_v \subset A[v]$ are subgroups such that H_v is of order p, L_v is of order p^2 and L_v is totally isotropic for the Weil pairing.

Let us spell out the definition of K^p-level structure. We may assume without loss of generality that S is connected, and we fix $\overline{s}$ a geometric point of S. The adelic Tate module $H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p})$ carries a symplectic Weil pairing

$$<,>_\lambda \colon H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \times H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \to H_1(\mathbf{G}_m|_{\overline{s}}, \mathbf{A}^{\infty,p})$$

or equivalently an F-linear symplectic pairing:

$$<,>_{1,\lambda} \colon H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \times H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \to H_1(\mathbf{G}_m|_{\overline{s}}, \mathbf{A}^{\infty,p}) \otimes F.$$

The level structure η is a K^p-orbit of pairs of isomorphisms (η_1, η_2), where (with $V = \mathcal{O}_F^4$ the standard symplectic space defined above):

(1) An $\mathcal{O}_F \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p}$-linear isomorphism of $\Pi_1(S, \overline{s})$-modules $\eta_1 : V \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p} \simeq H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p})$.
(2) An $\mathcal{O}_F \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p}$-linear isomorphism of $\Pi_1(S, \overline{s})$-modules $\eta_2 : F \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p} \simeq F \otimes_{\mathbf{Z}} H_1(\mathbf{G}_m|_{\overline{s}}, \mathbf{A}^{\infty,p})$.

We moreover impose that the following diagram is commutative:

$$\begin{tikzcd} V \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p} \times V \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p} \arrow[r, "\eta_1 \times \eta_1"] \arrow[d, "<,>_1"] & H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \times H_1(A|_{\overline{s}}, \mathbf{A}^{\infty,p}) \arrow[d, "<,>_{1,\lambda}"] \\ F \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p} \arrow[r, "\eta_2"] & F \otimes_{\mathbf{Z}} H_1(\mathbf{G}_m|_{\overline{s}}, \mathbf{A}^{\infty,p}) \end{tikzcd} \tag{3.3.4}$$

The action of an element $k \in K^p$ takes (η_1, η_2) to $(\eta_1 k, \nu(k)\eta_2)$.

Remark 3.3.5. The reader will observe that η_2 is uniquely determined by η_1, but we find it convenient to record it as part of the data for the sake of comparison to the PEL setting in Proposition 3.3.9 below.

A map between quintuples $(A, \iota, \lambda, \eta, \eta_p)$ and $(A', \iota', \lambda', \eta', \eta'_p)$ is an $\mathcal{O}_F$-linear prime to p quasi-isogeny (in the sense of [Lan13, Defn. 1.3.1.17]) $f : A \to A'$ such that

- $f^*\lambda = r\lambda'$ for a locally constant function $r : S \to \mathbf{Z}_{(p)}^{\times,+}$,
- $f(\eta_p) = \eta'_p$, and
- $H_1(f) \circ \eta = \eta'$.

This last condition means that η' is defined by $H_1(f) \circ \eta_1 = \eta'_1$ and $\eta'_2 = r^{-1}\eta_2$. Also, we have denoted $\mathbf{Z}_{(p)}^{\times,+} = \mathbf{Q}_{>0}^\times \cap \mathbf{Z}_{(p)}^\times$.

Remark 3.3.6. Note that we are allowing the similitude factor in the level structure to be in $\mathbf{A}^{\infty,p} \otimes_{\mathbf{Q}} F(1)$, but we only allow quasi-isogenies with similitude factor in $\mathbf{A}^{\infty,p}(1)$.

We denote by $\mathbf{Z}_{(p)}^{\times,+} \backslash (\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)(1)$ the set $\mathbf{Z}_{(p)}^{\times,+} \backslash (\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)$ equipped with the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ through the cyclotomic character $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \prod_{\ell \neq p} \mathbf{Z}_\ell^\times \hookrightarrow (\mathbf{A}^{\infty,p})^\times$. This action is unramified at p. It follows easily that $\mathbf{Z}_{(p)}^{\times,+} \backslash (\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)(1)$ is represented by an infinite disjoint union of finite étale schemes over $\operatorname{Spec} \mathbf{Z}_{(p)}$.

Remark 3.3.7. The group $\mathbf{Z}_{(p)}^{\times,+}$ acts freely on $(\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)$.

Remark 3.3.8. When $\nu(K^p) = (\mathcal{O}_F \otimes_{\mathbf{Z}} \prod_{\ell \neq p} \mathbf{Z}_\ell)^\times$, then the above Galois action is trivial and $\mathbf{Z}_{(p)}^{\times,+} \backslash (\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)(1)$ is simply an infinite disjoint union of copies of $\operatorname{Spec} \mathbf{Z}_{(p)}$.

There is a structural map $\Pi_0 : \mathbf{Y}_K \to \mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$ which associates to an object $(A,\iota,\lambda,\eta,\eta_p)$ of $\mathbf{Y}_K$ the class of $\eta_2(1)$ (where we are identifying $H_1(\mathbf{G}_m|_{\overline{s}},\mathbf{A}^{\infty,p})$ with $\mathbf{A}^{\infty,p}(1)$).

As we mentioned at the beginning of §3.2, the complex points of our integral models are not precisely the double coset spaces considered in §3.2, because our moduli problem only allows polarizations of degree prime to p. However, the difference amounts to throwing away some geometrically connected components, as the following result explains.

Proposition 3.3.9. *The groupoid $\mathbf{Y}_K$ is representable by a quasi-projective scheme $\Pi_0 : Y_K \to \mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$. The morphism Π_0 has geometrically connected fibres. Let $c \in \mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)$ and let $c : \operatorname{Spec}\mathbf{C} \to \mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$ be the associated morphism (for the usual choice of primitive roots of unity in $\mathbf{C}$). Let $Y_{K,c}$ be the fibre of Y_K over c. Then there is an isomorphism of analytic spaces $(Y_{K,c})^{an} = \Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$.*

Proof. This follows from the usual description of integral models of PEL type Shimura varieties; in the case of hyperspecial level this goes back to Kottwitz [Kot92], but for convenience we follow the notation of [Lan13]. To this end, we recall the description of these integral models for the usual Shimura varieties for G. We let $\widetilde{K} = \widetilde{K}^p\widetilde{K}_p$ denote a compact open subgroup of $G(\mathbf{A}^\infty)$, where $\widetilde{K}^p$ is a compact open subgroup of $G(\mathbf{A}^{\infty,p})$, and $\widetilde{K}_p$ is of one of the parahoric subgroups considered above.

Then we let $\mathbf{Y}^{G,\mathrm{Kott}}$ be the groupoid over $AFF/\mathbf{Z}_{(p)}$ whose fibre over $S \in \mathrm{Ob}(AFF/\mathbf{Z}_{(p)})$ is the category with objects $(A,\iota,\lambda,\tilde{\eta},\eta_p)$, where (A,ι,λ,η_p) is as in the definition of $\mathbf{Y}_K$ above, but now $\tilde{\eta}$ is given by a $\widetilde{K}^p$-orbit of pairs of isomorphisms $(\tilde{\eta}_1,\tilde{\eta}_2)$, consisting of:

(1) An $\mathcal{O}_F\otimes\mathbf{A}^{\infty,p}$-linear isomorphism of $\Pi_1(S,\overline{s})$-modules $\tilde{\eta}_1 : V\otimes_{\mathbf{Z}}\mathbf{A}^{\infty,p} \simeq H_1(A|_{\overline{s}},\mathbf{A}^{\infty,p})$.
(2) An $\mathbf{A}^{\infty,p}$-linear isomorphism of $\Pi_1(S,\overline{s})$-modules $\tilde{\eta}_2 : \mathbf{A}^{\infty,p} \simeq \mathbf{A}^{\infty,p}\otimes_{\mathbf{Z}} H_1(\mathbf{G}_m|_{\overline{s}},\mathbf{A}^{\infty,p})$.

We moreover impose that the following diagram is commutative:

(3.3.10)

$$\begin{tikzcd}
V\otimes_{\mathbf{Z}}\mathbf{A}^{\infty,p}\times V\otimes_{\mathbf{Z}}\mathbf{A}^{\infty,p} \arrow[r, "\tilde{\eta}_1\times\tilde{\eta}_1"] \arrow[d, "<,>"] & H_1(A|_{\overline{s}},\mathbf{A}^{\infty,p})\times H_1(A|_{\overline{s}},\mathbf{A}^{\infty,p}) \arrow[d, "<,>_\lambda"] \\
\mathbf{A}^{\infty,p} \arrow[r, "\tilde{\eta}_2"] & H_1(\mathbf{G}_m|_{\overline{s}},\mathbf{A}^{\infty,p})
\end{tikzcd}$$

A map between quintuples $(A,\iota,\lambda,\tilde{\eta},\eta_p)$ and $(A',\iota',\lambda',\tilde{\eta}',\eta_p')$ is an $\mathcal{O}_F$-linear prime to p quasi-isogeny $f : A \to A'$ such that

- $f^*\lambda = r\lambda'$ for a locally constant function $r : S \to \mathbf{Z}_{(p)}^{\times,+}$,
- $f(\eta_p) = \eta_p'$, and
- $H_1(f)\circ\tilde{\eta} = \tilde{\eta}'$.

It follows immediately from the definition that there is a natural isomorphism

$$\mathbf{Y}_K \cong \coprod_{g\in G(\mathbf{A}^{\infty,p})\backslash G_1(\mathbf{A}^{\infty,p})/K^p} \mathbf{Y}^{G,\mathrm{Kott}}_{gK^pg^{-1}\cap G(\mathbf{A}^{\infty,p})},$$

given by the maps

$$g : \mathbf{Y}^{G,\mathrm{Kott}}_{gK^pg^{-1}\cap G(\mathbf{A}^{\infty,p})} \to \mathbf{Y}_K$$

which are defined by

$$(A, \iota, \lambda, (\tilde{\eta}_1, \tilde{\eta}_2), \eta_p) \mapsto (A, \iota, \lambda, (\tilde{\eta}_1, \tilde{\eta}_2 \otimes_{\mathbf{Z}} \mathcal{O}_F)g, \eta_p).$$

(Indeed, one easily checks that this already gives a bijection of tuples before passing to isogeny classes, and that this bijection is compatible with isogenies.) The result now follows from [Kot92, §5, §8]. □

We now define an action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ (totally positive elements in $F^\times$ which are prime to p) on $\mathbf{Y}_K$, by scaling the polarization λ. Since this scales the λ-Weil pairing $\langle,\rangle_{1,\lambda}$, we see from (3.3.4) that it also scales η_2. Explicitly, $x \in (\mathcal{O}_F)^\times_{(p)}$ sends $(A, \iota, \lambda, (\eta_1, \eta_2), \eta_p)$ to $(A, \iota, x\lambda, (\eta_1, x\eta_2), \eta_p)$. By definition, the subgroup $\mathbf{Z}^{\times,+}_{(p)}$ acts trivially on $\mathbf{Y}_K$.

The group $(\mathcal{O}_F)^{\times,+}_{(p)}$ acts on the set of connected components $\Pi_0(Y_K)$. Since the cyclotomic character surjects onto $\prod_{l\neq p} \mathbf{Z}_l^\times$, the stabilizer of each connected component is

$$\mathcal{O}_F^{\times,+}(\Pi_0) := \left((\mathcal{O}_F)^{\times,+}_{(p)} \cap \mathbf{Z}^{\times,+}_{(p)} \nu(K^p) \prod_{\ell\neq p} \mathbf{Z}_\ell^\times \right) / \mathbf{Z}^{\times,+}_{(p)},$$

which we can and do naturally identify with

$$\mathcal{O}_F^{\times,+} \cap \nu(K^p) \prod_{\ell\neq p} \mathbf{Z}_\ell^\times.$$

Remark 3.3.11. If $\nu(K^p) = (\mathcal{O}_F \otimes_{\mathbf{Z}} \prod_{\ell\neq p} \mathbf{Z}_\ell)^\times$, then $\mathcal{O}_F^{\times,+}(\Pi_0) = \mathcal{O}_F^{\times,+}$.

The subgroup $\mathcal{O}_F^{\times,+}(\nu(K^p)) := \mathcal{O}_F^{\times,+} \cap \nu(K^p)$ acts trivially on each connected component of $\Pi_0(Y_K)$. The quotient stack of connected components is

$$[((\mathcal{O}_F)^{\times,+}_{(p)}/\mathbf{Z}^{\times,+}_{(p)})\backslash\big(\mathbf{Z}^{\times,+}_{(p)}\backslash(\mathbf{A}^{\infty,p} \otimes F)^\times/\nu(K^p)(1)\big)].$$

It admits a coarse moduli space $(\mathcal{O}_F)^{\times,+}_{(p)}\backslash(\mathbf{A}^{\infty,p} \otimes F)^\times/\nu(K^p)(1)$ which is a finite étale covering of $\operatorname{Spec} \mathbf{Z}_{(p)}$.

We now take the quotient stack

$$\mathbf{Y}_K^{G_1} := [Y_K/((\mathcal{O}_F)^{\times,+}_{(p)}/\mathbf{Z}^{\times,+}_{(p)})].$$

This is the "Shimura stack" associated to G_1 and the level K.

Let us define

$$\mathcal{O}_F^{\times,+}(K^p) = \{x^2 \mid x \in \mathcal{O}_F^\times \cap K^p\},$$

where $\mathcal{O}_F^\times$ is thought of inside $G_1(\mathbf{A}_f^{(p)})$ as a subgroup of the scalar matrices. The multiplier of the scalar matrix given by x is x^2, and hence the multiplier of $\mathcal{O}_F^{\times,+}(K^p)$ lands inside $\nu(K^p)$, and hence $\mathcal{O}_F^{\times,+}(K^p)$ is a finite index subgroup of $\mathcal{O}_F^{\times,+}(\nu(K^p))$ and of $\mathcal{O}_F^{\times,+}(\Pi_0)$.

Lemma 3.3.12. *The restriction of the action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ on Y_K to $\mathcal{O}_F^{\times,+}(K^p)$ is trivial. More precisely, there is a canonical natural transformation going from the action of $\mathcal{O}_F^{\times,+}(K^p)$ on $\mathbf{Y}_K$ to the trivial action of $\mathcal{O}_F^{\times,+}(K^p)$ on $\mathbf{Y}_K$.*

Proof. Let $x^2 \in \mathcal{O}_F^{\times,+}(K^p)$ for a unique $x \in \mathcal{O}_F^{\times,+} \cap K^p$. The action of x^2 sends $(A, \iota, \lambda, \eta, \eta_p)$ to $(A, \iota, x^2\lambda, \eta, \eta_p)$ (note that since $x \in K^p$, and η is by definition a K^p-orbit, the action of x^2 on η is trivial). On the other hand multiplication by $x^{-1} : A \to A$ provides a map $(A, \iota, \lambda, \eta, \eta_p) \to (A, \iota, x^2\lambda, \eta, \eta_p)$ in the groupoid $\mathbf{Y}_K$. This provides the natural transformation from the action of x^2 obtained from the action of $(\mathcal{O}_F)_{(p)}^{\times,+}$ to the trivial action. □

Lemma 3.3.13. *For any geometric point $x \in Y_K$, the stabilizer of x for the action of $(\mathcal{O}_F)_{(p)}^{\times,+}$ is $\mathcal{O}_F^{\times,+}(K^p)$.*

Proof. By Lemma 3.3.12, $\mathcal{O}_F^{\times,+}(K^p)$ is contained in the stabilizer of any $x = (A, \iota, \lambda, \eta)$. Let $\epsilon \in (\mathcal{O}_F)_{(p)}^{\times,+}$ and assume that there is a morphism

$$f : (A, \iota, \lambda, (\eta_1, \eta_2), \eta_p) \to (A, \iota, \epsilon\lambda, (\eta_1, \epsilon\eta_2), \eta_p)$$

in the groupoid $\mathbf{Y}_K$. We need to show that $f \in \mathcal{O}_F^\times \cap K^p$. Since f respects η_1, it follows from [Lan13, Lem. 1.3.5.2] that f is an automorphism of A (and not just a quasi-isogeny).

The polarization λ induces an involution $x \mapsto \bar{x}$ on $F(f)$, and we consider the automorphism $\alpha = f\bar{f}^{-1}$ of A. It stabilizes the polarization: $\alpha^*\lambda = \lambda\alpha\bar{\alpha} = \lambda$. It also stabilizes the level structure: $\bar{f}$ acts like the adjoint of f on $H_1(A, \mathbf{A}^{\infty,p})$. Since K^p is neat, this implies that $\alpha = 1$; indeed, all the eigenvalues of α are roots of unity, because they are algebraic numbers all of whose conjugates have absolute value 1. It follows that $f = \bar{f}$, and $f^2 = f\bar{f} = \epsilon$. Since f is an automorphism, it follows that $\epsilon \in \mathcal{O}_F^\times$. Hence it suffices to show that $f \in F$, since we then have $\epsilon \in \mathcal{O}_F^{\times,+}(K^p)$.

Assume first that A is simple, so that $\mathrm{End}(A)_{\mathbf{Q}}$ is a division algebra and $F(f) \subset \mathrm{End}(A)_{\mathbf{Q}}$ is a commutative field on which the Rosati involution $x \mapsto \bar{x}$ is complex conjugation. Since $f = \bar{f}$ and $f^2 = \epsilon$, $F(f)$ is a totally real extension of F of degree at most 2. If $F(f) = F$, we are done. Otherwise $F(f)$ is a quadratic extension of F. The level structure η provides a K^p-orbit of isomorphisms $H_1(A, \mathbf{A}^{\infty,p}) \simeq V \otimes \mathbf{A}^{\infty,p}$, and the element f acts via some conjugate of

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

and has eigenvalues in $\bar{F}$: $\{\sqrt{\epsilon}, -\sqrt{\epsilon}\}$ with multiplicity two. By neatness, no conjugate of this matrix is in K^p, a contradiction.

We now assume that A is not simple. It is easy to see (using the $\mathcal{O}_F$-action) that the only possibility is that A is isogenous to $A_1 \times A_2$ where A_1 and A_2 are two abelian schemes of dimension $[F : \mathbf{Q}]$ with $F \subset \mathrm{End}(A_i)_{\mathbf{Q}}$. If A_1 and A_2 are not isogenous, then $\mathrm{End}(A)_{\mathbf{Q}} = \mathrm{End}(A_1)_{\mathbf{Q}} \times \mathrm{End}(A_2)_{\mathbf{Q}}$. Moreover, $F(f)$ is a commutative subalgebra of $\mathrm{End}(A_1)_{\mathbf{Q}} \times \mathrm{End}(A_2)_{\mathbf{Q}}$ and is therefore included in a product of fields $F_1 \times F_2$ where F_i is either F or a CM extension of F. Since $f = \overline{f}$, we see that $f = (f_1, f_2) \in F \times F$ and that $f^2 = (f_1^2, f_2^2) = \epsilon$. So either $f_1 = f_2$, and we are done, or $f_1 = -f_2$; but this second case is again prohibited by neatness.

Lastly, we assume that A is isogenous to A_1^2. Then $\mathrm{End}(A)_{\mathbf{Q}} \simeq M_2(\mathrm{End}(A_1)_{\mathbf{Q}})$ and $F(f)$ is a commutative subalgebra, therefore included in $M_2(E)$ where E is

either F or a CM extension of F. Writing $f = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(E)$, we have $\epsilon = f^2 = \begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & bc+d^2 \end{pmatrix}$. If $a+d=0$, the matrix of f has eigenvalues $\{\sqrt{\epsilon}, -\sqrt{\epsilon}\}$ and this is again impossible by neatness. We deduce that $a+d \neq 0$, so that $b=c=0$ and $a=d=\sqrt{\epsilon}$ or $a=d=-\sqrt{\epsilon}$. Since $\overline{f}=f$ and the Rosati involution induces the complex conjugation on E, we deduce that $\sqrt{\epsilon} \in F$ and that $f \in F$, as required. □

We write

$$\Delta = (\mathcal{O}_F)^{\times,+}_{(p)}/\mathcal{O}_F^{\times,+}(K^p), \tag{3.3.14}$$

$$\Delta(\Pi_0) = \mathcal{O}_F^{\times,+}(\Pi_0)/\mathcal{O}_F^{\times,+}(K^p), \tag{3.3.15}$$

$$\Delta(K^p) = \mathcal{O}_F^{\times,+}(\nu(K^p))/\mathcal{O}_F^{\times,+}(K^p). \tag{3.3.16}$$

These last two groups are finite groups. Let us set $Y_K^{G_1} = \Delta\backslash Y_K$. This last quotient exists as a scheme. Indeed, Δ permutes the connected components of Y_K and the stabilizer of any connected component is a finite group $\Delta(\Pi_0)$, while the stabilizer of any geometrically connected component if $\Delta(K^p)$. Moreover, the action of Δ can be lifted to an action on an ample line bundle on Y_K (for instance the tensor product of the line bundles $\det(\Omega^1_{(A/C)/Y_K})$ where C runs over all subgroups $C = \prod_{v|p} C_v$ where for each $v \mid p$, C_v is either 1 or whichever of H_v, L_v exist as part of the level structure, see [Lan16, §6]). The group $\Delta(\Pi_0)$ acts without fixed points by Lemma 3.3.13. The following proposition then follows immediately from Proposition 3.3.9 and Lemma 3.2.3.

Proposition 3.3.17. *There is a canonical map $\mathbf{Y}_K^{G_1} \to Y_K^{G_1}$, and $Y_K^{G_1}$ is the coarse moduli of $\mathbf{Y}_K^{G_1}$. There is a quasi-projective morphism $\Pi_0 : Y_K^{G_1} \to (\mathcal{O}_F)^{\times,+}_{(p)}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$ with geometrically connected fibres. Moreover, the map $Y_K \to Y_K^{G_1}$ is étale and surjective.*

Let $c \in \mathbf{Z}^{\times,+}_{(p)}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)$ and let

$$c : \operatorname{Spec} \mathbf{C} \to \mathbf{Z}^{\times,+}_{(p)}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$$

be the associated morphism (for the usual choice of primitive roots of unity in $\mathbf{C}$). Let $Y_{K,c}$ be the fibre of Y_K over c and let $Y_{K,c}^{G_1}$ be the fibre of $Y_K^{G_1}$ over c. Then there is a commutative diagram of analytic spaces where the horizontal maps are isomorphisms and the vertical maps are finite étale with groups $\Delta(K^p) = \Delta(K)$:

$$\begin{array}{ccc} (Y_{K,c})^{an} & \longrightarrow & \Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})} \\ \downarrow & & \downarrow \\ (Y_{K,c}^{G_1})^{an} & \longrightarrow & \Gamma_1(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})} \end{array}$$

3.4. **Local models.** We now recall some basic results about local models for GSp_4; the cases that we need essentially go back to [dJ93]. Continue to let K be a reasonable level structure. For each place $v|p$, we let $M^{\mathrm{loc}}_{K_v}$ be the moduli space over $\mathcal{O}_{F_v}$ of chains of lattices corresponding to K_v; so for example $M^{\mathrm{loc}}_{\mathrm{Par}(v)}$ is the moduli space of totally isotropic direct factors of $V_1 \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v}$ of rank 2. We write

$M^{\mathrm{loc}}_{K_p} := \times_{v|p} M^{\mathrm{loc}}_{K_v}$. Then by the results of [RZ96, §6], each geometric point of the special fibre of Y^G_K has an étale neighbourhood which is isomorphic to an étale neighbourhood of a geometric point in $M^{\mathrm{loc}}_{K_p}$. (The description of the local model in [RZ96, §6] is in terms of chains of $\mathcal{O}_F \otimes \mathbf{Z}_p$-lattices, but this description can be immediately rewritten in terms of products over the places $v|p$ of chains of $\mathcal{O}_{F_v}$-lattices.)

Proposition 3.4.1. *The scheme $\mathbf{Y}_K$ is flat over $\operatorname{Spec} \mathbf{Z}_{(p)}$, normal, and a local complete intersection (so in particular Cohen–Macaulay) of pure relative dimension $3[F:\mathbf{Q}]$. If $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ for all $v|p$, then it is smooth, while in general it is smooth away from codimension 2.*

Proof. Note that normality follows from being smooth away from codimension 2 and Cohen–Macaulay. The properties of being flat and a local complete intersection over $\operatorname{Spec} \mathbf{Z}_{(p)}$, and of being smooth, or smooth away from codimension 2, can all be checked étale locally ([Sta13, Tag 03E7,Tag 04R3,Tag 06C3]). Furthermore, these properties are all preserved by taking products. It therefore suffices to show that they hold for the local models $M^{\mathrm{loc}}_{K_v}$. This has already been carried out in the literature: the case that $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ is trivial, and the cases that $K_v = \mathrm{Kli}(v)$, $\mathrm{Si}(v)$ or $\mathrm{Iw}(v)$ are covered in [Til06b, §2]. In the case $K_v = \mathrm{Par}(v)$ see [Yu11, Prop. 2.5, Thm. 2.11]. □

Corollary 3.4.2. *The scheme $Y^{G_1}_K$ is normal, flat over $\operatorname{Spec} \mathbf{Z}_{(p)}$, and a local complete intersection.*

Proof. Since $Y_K \to Y^{G_1}_K$ is an étale surjection by Proposition 3.3.17, this is immediate from Proposition 3.4.1. □

3.5. **Compactifications.** In this section, we state results on the existence of toroidal compactifications. Toroidal compactifications depend on some combinatorial data which we first explain. We will follow closely the presentation of [Pin90] and [HLTT16], see in particular [HLTT16, §5.2] (that this presentation is equivalent to Lan's presentation is explained in [HLTT16, App. B]).

In this section, we write V_F for $V \otimes_{\mathcal{O}_F} F$. Let $\mathfrak{C}$ be the set of totally isotropic F-subspaces $W \subset V_F$. For all $W \in \mathfrak{C}$, consider the $F \otimes \mathbf{R}$-module of $\mathbf{Q}$-bilinear forms

$$\phi : V_F/W^\perp \times V_F/W^\perp \to \mathbf{R}$$

which satisfy $\phi(\lambda x, y) = \phi(x, \lambda y)$ for all $\lambda \in F$, $x, y \in V_F/W^\perp$. Let $C(V_F/W^\perp)$ be the cone inside this $\mathbf{R}$-vector space given by those forms which are positive semidefinite and whose radical is defined over F. Let $\mathcal{C}$ be the conical complex which is the quotient of $\coprod_{W \in \mathfrak{C}} C(V_F/W^\perp)$ by the equivalence relation induced by the inclusions $C(V_F/W^\perp) \subset C(V_F/Z^\perp)$ for $W \subset Z$.

A *non-degenerate rational polyhedral cone* of $\mathcal{C} \times G_1(\mathbf{A}^\infty)$ is a subset contained in $\mathcal{C}(V_F/W^\perp) \times \{\gamma\}$ for some (W, γ) which is of the form $\sum_{i=1}^k \mathbf{R}_{>0} s_i$ for elements $s_i : V_F/W^\perp \times V_F/W^\perp \to \mathbf{Q}$.

A *rational polyhedral cone decomposition* Σ of $\mathcal{C} \times G_1(\mathbf{A}^\infty)$ is a partition $\mathcal{C} \times G_1(\mathbf{A}^\infty) = \coprod_{\sigma \in \Sigma} \sigma$ by non-degenerate rational polyhedral cones σ such that the closure of each cone is a union of cones.

Let $W \in \mathfrak{C}$. We let P_W be the parabolic subgroup of G_1 which is the stabilizer of W. Let us denote by $M_{W,l}$ the group of F-linear automorphisms of $V_F/W^\perp$. We

also denote by $M_{W,h}$ the group of symplectic similitudes of $W^\perp/W$ (so that this group is isomorphic to $\mathrm{Res}_{F/\mathbf{Q}}\mathrm{GSp}_{4-2\dim W}$, and in particular is non-trivial even when $\dim W = 2$). The group $M_W = M_{W,h} \times M_{W,l}$ is the Levi quotient of P_W. We have a surjective map $P_W \to M_{W,l}$, and we denote by $P_{W,h}$ its kernel. There is a surjective map $P_{W,h} \to M_{W,h}$.

The group $G_1(\mathbf{Q})^+$ acts on $\mathfrak{C}$ and also on $\mathcal{C}$. Let $W \in \mathfrak{C}$, let $\gamma \in G_1(\mathbf{Q})^+ \cap P_W$ and $\phi \in C(V_F/W^\perp)$. Let γ_l be the projection of γ in $M_{W,l}$. Then we set $\gamma\phi(x,y) = \nu(\gamma)\phi(\gamma_l.x, \gamma_l.y)$.

The set $\mathcal{C} \times G_1(\mathbf{A}^\infty)$ carries a diagonal left action of $G_1(\mathbf{Q})$ and left and right actions of $G_1(\mathbf{A}^\infty)$ (by left and right multiplication on the second factor). For any compact open subgroup $K \subset G_1(\mathbf{A}^\infty)$, a rational polyhedral cone decomposition Σ is K-equivariant if for all $h \in G_1(\mathbf{Q}), k \in K$ and $\sigma \in \Sigma$, $h.\sigma.k \in \Sigma$.

For any compact open subgroup $K \subset G_1(\mathbf{A}^\infty)$ we say that a rational polyhedral cone decomposition Σ of $\mathcal{C} \times G_1(\mathbf{A}^\infty)$ is K-admissible if:

(1) The decomposition is K-equivariant.
(2) For all $\sigma \subset C(V_F/W^\perp) \times \{\gamma\}$, and all $p \in P_{W,h}(\mathbf{A}^\infty)$, we have $p.\sigma \in \Sigma$.
(3) For all cones σ, let $W \in \mathfrak{C}$ be such that $\sigma \subset \mathcal{C}(V_F/W^\perp)$ is in the interior of $\mathcal{C}(V_F/W^\perp)$. Then if there are $p \in P_{W,h}(\mathbf{A}^\infty)$, $u \in K$ and $h \in G_1(\mathbf{Q})$ satisfying $\sigma \cap hp\sigma u \neq \emptyset$, then in fact $h \in P_{W,h}(\mathbf{A}^\infty)$.
(4) $G_1(\mathbf{Q})\backslash\Sigma/K$ is finite.

There exist K-admissible rational polyhedral cone decompositions. Any two K-admissible rational polyhedral cone decompositions can be refined by a third one.

If $L_W \subset \mathrm{Hom}_{\mathbf{Q}}(\mathrm{Sym}^2_F V_F/W^\perp, \mathbf{Q})$ is a lattice, then a cone

$$\sigma \subset \mathrm{Hom}_{\mathbf{Q}}(\mathrm{Sym}^2_{\mathcal{O}_F} V_F/W^\perp, \mathbf{Q})$$

is said to be smooth with respect to L_W if the s_i can be taken to be part of a basis of L_W. Assume that for all $(W,\gamma) \in \mathfrak{C} \times G_1(\mathbf{A}^\infty)$ we have lattices

$$L_{W,\gamma} \subset \mathrm{Hom}_{\mathbf{Q}}(\mathrm{Sym}^2_F V_F/W^\perp, \mathbf{Q}).$$

We say that a rational polyhedral cone decomposition Σ is smooth with respect to these lattices if each cone $\sigma \in \Sigma$ is smooth.

We now assume that $K = K^p K_p$ is a reasonable compact open subgroup. We choose a lattice $V' \subset V_F$ with the property that K^p stabilizes $V' \otimes_{\mathbf{Z}} \mathbf{A}^{\infty,p}$ and that $V' \otimes \mathcal{O}_{F_v} = V \otimes \mathcal{O}_{F_v}$ for all places $v|p$ such that $K_v \neq \mathrm{Par}(v)$ and $V' \otimes \mathcal{O}_{F_v} = V_3$ for all places v such that $K_v = \mathrm{Par}(v)$.

Then $(\mathcal{O}_F, V', \langle . \rangle)$ defines an integral PEL datum and $K \subset G_1'(\hat{\mathbf{Z}})$ where G_1' is the group scheme over $\mathrm{Spec}\,\mathbf{Z}$ of symplectic similitudes of V'.

The theory of toroidal compactification associates a lattice $L_{W,K,\gamma} \subset C(V_F/W^\perp)$ to this integral PEL datum, compact open K, $W \in \mathfrak{C}$ and $\gamma \in G_1(\mathbf{A}^\infty)$ (see [Lan13, §5.3] and [Lan16, §3]). The K-admissible rational polyhedral cone decompositions which satisfy the following extra properties form a cofinal subset of the set of all K-admissible rational polyhedral cone decompositions:

(1) The decomposition is projective (in the sense of [AMRT10]).
(2) The decomposition is smooth with respect to the lattices $L_{W,K,\gamma}$.

In the rest of the paper, we will consider K-admissible rational polyhedral cone decompositions which satisfy these extra properties unless explicitly stated.

Theorem 3.5.1.

(1) *Let Σ be a K-admissible polyhedral cone decomposition which is projective. There is a toroidal compactification $X_{K,\Sigma}$ of Y_K. It has a stratification indexed by $(G(\mathbf{Q})^+ \cap K_p)\backslash\Sigma/K^p = G(\mathbf{Q})^+\backslash\Sigma/K$. The boundary is the reduced complement of Y_K in $X_{K,\Sigma}$. This is a relative Cartier divisor denoted by $D_{K,\Sigma}$.*
(2) *The universal abelian scheme $A \to Y_K$ extends to a semi-abelian scheme $A \to X_{K,\Sigma}$.*
(3) *If Σ' is a refinement of Σ, then there are projective maps $\pi_{\Sigma',\Sigma} : X_{K,\Sigma'} \to X_{K,\Sigma}$, and $(\mathrm{R}\pi_{\Sigma',\Sigma})_*\mathcal{O}_{X_{K,\Sigma'}} = \mathcal{O}_{X_{K,\Sigma}}$. Let $\mathcal{I}_{X_{K,\Sigma}}$ and $\mathcal{I}_{X_{K,\Sigma'}}$ be the invertible sheaves of the boundary in $X_{K,\Sigma}$ and $X_{K,\Sigma'}$. Then $\pi^*_{\Sigma',\Sigma}\mathcal{I}_{X_{K,\Sigma}} = \mathcal{I}_{X_{K,\Sigma'}}$ and $(\mathrm{R}\pi_{\Sigma',\Sigma})_*\mathcal{I}_{X_{K,\Sigma'}} = \mathcal{I}_{X_{K,\Sigma}}$.*
(4) *Suppose that K is reasonable (in the sense of Definition 3.3.3). Then the toroidal compactification $X_{K,\Sigma}$ is flat over $\operatorname{Spec}\mathbf{Z}_{(p)}$, normal, and Cohen-Macaulay. If Σ is smooth, then $X_{K,\Sigma} \to \operatorname{Spec}\mathbf{Z}_{(p)}$ is further a local complete intersection. Finally if $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ for all $v|p$ and Σ is smooth then $X_{K,\Sigma} \to \operatorname{Spec}\mathbf{Z}_{(p)}$ is smooth.*

Proof. This follows from [Lan17, Thm. 6.1]. We simply need to specify the choices we made to construct the toroidal compactification by normalization (see [Lan16, §2]). In the first case that $K_p = \mathrm{G}_1(\mathbf{Z}_p)$ (the nice case: no level at p, prime to p polarization), the compactification is constructed in [Lan13]. In the second case that $K_p = \prod_{v|p} K_v$ where $K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v)\}$, the compactification can be constructed as a closed subscheme of some toroidal compactification of a Siegel modular variety with a prime to p polarization (Zarhin's trick) (and possibly performing again a blow up or a blow down at the boundary as explained in [Lan17]). In the general case where we have a parahoric level structure, we consider all possible degeneration maps $Y_K \to \prod_{K'_p} Y_{K^pK'_p}$ where $K_p \to K'_p$ and $K'_p = \prod_{v|p} K'_v$ with $K'_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v)\}$ and obtain the toroidal compactification as a closed subscheme of the product of the toroidal compactifications of the $Y_{K^pK'_p}$ (and possibly performing again a blow up or a blow down at the boundary as explained in [Lan17]).

Now, everything apart from (4) is immediate, while (4) follows from Proposition 3.5.4 together with the explicit description of the formal completions along boundary strata given in [Lan17, Thm. 6.1 (4)]. □

We also need to consider the action of the group $\mathcal{O}^{\times,+}_{F,(p)}$. Recall that we defined a quotient Δ of this group in (3.3.15).

Lemma 3.5.2. *The action of $\mathcal{O}^{\times,+}_{F,(p)}$ on Y_K extends to $X_{K,\Sigma}$ and factors through Δ.*

Proof. It is possible to prove this directly by looking at the construction of the toroidal compactification and the boundary charts. We will instead give a simpler indirect argument. Since $X_{K,\Sigma}$ is normal, it follows that $X_{K,\Sigma}$ is the normalization of Y_K in $X_{K,\Sigma} \times \operatorname{Spec}\mathbf{C}$. It is therefore sufficient to show that the action extends over $\mathbf{C}$.

We can now use [AMRT10]. Let $c \in G_1(\mathbf{A}_f)$. By Proposition 3.3.9, the analytification of the component $Y_{K,c} \subset Y_K \times \operatorname{Spec}\mathbf{C}$ corresponding to c is $\Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$, and we need to show that the group $\Delta(K)$ (which is the subgroup of Δ acting trivially on the geometrically connected components) acts on the compactification of $\Gamma(c,K)\backslash\mathcal{H}^{\mathrm{Hom}(F,\mathbf{R})}$. By the main results of [AMRT10], our choice of Σ provides a

partial compactification $\mathcal{H}_\Sigma^{\mathrm{Hom}(F,\mathbf{R})}$ which carries an action of $\Gamma(K,c)$. The component of $(X_{K,\Sigma}\times \operatorname{Spec}\mathbf{C})^{an}$ corresponding to c is isomorphic to $\Gamma(c,K)\backslash\mathcal{H}_\Sigma^{\mathrm{Hom}(F,\mathbf{R})}$. This space still carries an action of $\Gamma_1(c,K)/\Gamma(c,K)$, which is what we claimed. □

Lemma 3.5.3. *The action of Δ on $X_{K,\Sigma}$ is free.*

Proof. Over Y_K, this is the content of Lemma 3.3.13. We claim that the action of Δ is free on the set of non-trivial strata in $X_{K,\Sigma}$. This set is simply $G^+(\mathbf{Q})\backslash\big(\Sigma\backslash\{0\}\times\mathbf{G}(\mathbf{A}^\infty)\big)/K$. Let $c\in G_1(\mathbf{A}^\infty)$, $\Gamma(c,K)=G(\mathbf{Q})^+\cap cKc^{-1}$ and $\Gamma_1(c,K)=G_1(\mathbf{Q})^+\cap cKc^{-1}$. Let Σ_c be the restriction of Σ to $\mathcal{C}\times\{c\}$. We need to show that the stabilizer of $\Gamma_1(c,K)$ acting on $\Sigma_c\setminus\{0\}$ is included in $\Gamma(c,K)$. This will imply that the group $\Delta(K)$ acts freely on $\Gamma(c,K)\backslash(\Sigma_c\setminus\{0\})$.

Let $W\in\mathfrak{C}\setminus\{0\}$. We denote by $\Gamma_W(c,K)$ and $\Gamma_{1,W}(c,K)$ the intersections of P_W with $\Gamma(c,K)$ and $\Gamma_1(c,K)$ respectively. Let $\sigma\subset C(V_F/W^\perp)\times\{c\}$ in the interior. By our assumption on the cone decomposition, if an element $\gamma\in\Gamma_{1,W}(c,K)$ stabilizes σ, then its linear part γ_l is trivial. We need to see that $\nu(\gamma)$ is trivial. It is easy to see that we can find an element $\gamma'\in\Gamma_W(c,K)$ and $n\in\mathbf{Z}_{\geq 0}$ such that $\nu(\gamma)^n\phi=\gamma'.\phi$ for all $\phi\in C(V_F/W^\perp)$ (it follows from the very definition of the action that the image of $\Gamma_W(c,K)$ in the space of automorphisms of $C(V_F/W^\perp)$ contains a finite index subgroup of $\mathcal{O}_F^{\times,+}$). We deduce that γ' stabilizes σ and therefore $\gamma'_l=1$, so that $\nu(\gamma)^n=1$ and $\nu(\gamma)=1$ since $\mathcal{O}_F^{\times,+}$ is torsion free. □

We form the quotient of $X_{K,\Sigma}$ by the action of $\mathcal{O}_{F,(p)}^{\times,+}$. This quotient exists because, on a given connected component of $X_{K,\Sigma}$, this is the quotient by a finite group, and the component is projective because Σ is a projective cone decomposition. We shall call such a quotient a toroidal compactification $X_{K,\Sigma}^{G_1}$ of $Y_K^{G_1}$. We summarize our findings in the following proposition:

Proposition 3.5.4. *The space $X_{K,\Sigma}^{G_1}$ has a stratification indexed by $G_1(\mathbf{Q})^+\backslash\Sigma/K$. The map $X_{K,\Sigma}\to X_{K,\Sigma}^{G_1}$ is étale and surjective. If K is reasonable, then $X_{K,\Sigma}^{G_1}$ is a flat local complete intersection over* $\operatorname{Spec}\mathbf{Z}_{(p)}$*, and is normal.*

If not necessary, we drop the subscripts K or Σ and simply write X. We denote the boundary divisor by D.

3.6. **Functorialities.** We now briefly discuss some functorial maps between Shimura varieties at different levels, which we will make use of when we discuss Hecke operators in §3.8. All of the functorialities that we consider here extend to the toroidal compactification for suitable choices of cone decompositions, so we confine our discussions to the interior.

3.6.1. *Change of level away from p.* Let $K=K^pK_p$ and $K'=(K^p)'K_p$ be two compact open subgroups of $G_1(\mathbf{A}^\infty)$ such that $K\subset K'$. Then we have finite étale maps $Y_K\to Y_{K'}$ and $Y_K^{G_1}\to Y_{K'}^{G_1}$, given by "forgetting the level structure"; that is, by replacing the K^p-orbit by the corresponding $(K^p)'$-orbit.

3.6.2. *Action of the group $G_1(\mathbf{A}^{\infty,p})$.* Let $g\in G_1(\mathbf{A}^{\infty,p})$. Then we can define an isomorphism

$$[g]:Y_K\to Y_{g^{-1}Kg}$$

by sending an object $(A,\iota,\lambda,\eta,\eta_p)$ of $\mathbf{Y}_K$ to $(A,\iota,\lambda,\eta\circ g,\eta_p)$, which is immediately seen to be an object of $\mathbf{Y}_{g^{-1}Kg}$.

We deduce isomorphisms $[g] : Y_K^{G_1} \to Y_{g^{-1}Kg}^{G_1}$.

3.6.3. *Change of level at p: Klingen type correspondences.* We now fix K^p and a place w above p. We let $K_p = \prod_{v|p} K_v \subset G_1(\mathbf{Z}_p)$ be a reasonable compact open such that $K_w = \mathrm{GSp}_4(\mathcal{O}_{F_w})$. We let $K'_p = \prod_{v \neq w} K_v \times \mathrm{Kli}(w)$ be another reasonable level structure at p and let $K''_p = \prod_{v \neq w} K_v \times \mathrm{Par}(w)$. Set $K = K^p K_p$, $K' = K^p K'_p$ and $K'' = K^p K''_p$.

Lemma 3.6.4. *There are natural proper surjective, generically finite étale forgetful maps $p_1 : Y_{K'} \to Y_K$ and $p_1 : Y_{K'}^{G_1} \to Y_K^{G_1}$.*

Proof. We simply forget the level structure H_w at w. □

We now choose once and for all an element $x_w \in F^{\times,+}$ which is a uniformizing element in F_w and a unit in F_v for all $v \neq w$ above p. This element is well defined up to multiplication by an element of $(\mathcal{O}_F)^{\times,+}_{(p)}$.

Lemma 3.6.5. *There is a proper, surjective, generically finite étale map $p_2 : Y_{K'} \to Y_{K''}$ depending on x_w and sending A to $A/H_w^\perp$. It induces a canonical map $p_2 : Y_{K'}^{G_1} \to Y_{K''}^{G_1}$.*

Proof. This map is defined to take an object $(A, \iota, \lambda, \eta, \eta_p)$ of $\mathbf{Y}_{K'}$ to the object $(A', \iota', \lambda', \eta', \eta'_p) \in \mathbf{Y}_{K''}$ defined as follows:

- $A' = A/H_w^\perp$, where $H_w^\perp \subset A[w]$ is an order p^3 group scheme, the orthogonal complement of H_w for the Weil pairing. Write $\pi : A \to A'$ for the natural isogeny.
- $\iota'(x) = \pi \circ \iota(x) \circ \pi^{-1}$,
- The quasi-polarization λ' is obtained by descending the quasi-polarization $x_w^2.\lambda$ from A to A'.
- $\eta' = \pi \circ \eta$.
- η'_p is the data of level structures at places $v \neq w$ above p deduced from η_p by the isomorphisms $\pi : A[v] \to A'[v]$.

The ambiguity in the choice of x_w disappears when we pass to the quotient stacks by the action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ and pass to the associated coarse moduli. □

Remark 3.6.6. There is another map $Y_{K'} \to Y_{K''}$ obtained by sending an abelian surface A to A/H_w; however, we will not need to make use of this map.

3.6.7. *Change of level at p: Siegel type correspondences.* We now fix K^p and a place w above p. We let $K_p = \prod_{v|p} K_v \subset G_1(\mathbf{Z}_p)$ be a reasonable compact open such that $K_w = \mathrm{GSp}_4(\mathcal{O}_{F_w})$ (resp. $\mathrm{Kli}(w)$). We let $K'_p = \prod_{v \neq w} K_v \times \mathrm{Si}(w)$ be another reasonable level structure at p (resp. $K'_p = \prod_{v \neq w} K_v \times \mathrm{Iw}(w)$). Set $K = K^p K_p$, $K' = K^p K'_p$.

Remark 3.6.8 (Warning). Note that the use of K and K' (and p_2) in this section (§3.6.7) differs from that in the previous section (§3.6.3). Thus the reader should be careful when these maps are used to note whether we are in the Klingen or Siegel setting (we indicate in any ambiguous context by giving references to the corresponding section). We made this choice since otherwise the number of required subscripts would become excessively cumbersome.

Lemma 3.6.9. *There are natural forgetful maps $p_1 : Y_{K'} \to Y_K$ and $p_1 : Y_{K'}^{G_1} \to Y_K^{G_1}$ which are surjective and generically finite.*

Proof. We simply forget the level structure L_w at w. □

Recall that we have chosen an element $x_w \in F^{\times,+}$ which is a uniformizing element in F_w and a unit in F_v for all $v \neq w$ above p.

Lemma 3.6.10. *There is a map $p_2 : Y_{K'} \to Y_K$ depending on x_w. It induces a canonical map $p_2 : Y_{K'}^{G_1} \to Y_K^{G_1}$.*

Proof. We take an object $(A, \iota, \lambda, \eta, \eta_p)$ of $\mathbf{Y}_{K'}$. We define $(A', \iota', \lambda', \eta', \eta'_p) \in \mathbf{Y}_K$ as follows:

- $A' = A/L_w$, call $\pi : A \to A'$ the isogeny.
- $\iota'(x) = \pi \circ \iota(x) \circ \pi^{-1}$.
- The quasi-polarization λ' is obtained by descending the quasi-polarization $x_w\lambda$ from A to A'.
- $\eta' = \pi \circ \eta$.
- η'_p is a data of level structures at places $v \neq w$ above p deduced from η_p by the isomorphisms $\pi : A[v] \to A'[v]$.
- In the case $K_w = \mathrm{Kli}(w)$, we define $H'_w = H_w^{\perp}/L_w \subset A'[w]$.

The ambiguity in the choice of x_w disappears when we pass to the quotient stacks by the action of $(\mathcal{O}_F)_{(p)}^{\times,+}$ and pass to the associated coarse moduli. □

3.7. **Automorphic vector bundles.** We now work over $\mathbf{Z}_p$, and assume from now on that p splits completely in F. We let S_p be the set of places of F above p. We have a decomposition $\mathcal{O}_F \otimes_{\mathbf{Z}} \mathbf{Z}_p = \prod_{v|p} \mathbf{Z}_p$. We also denote by $v : \mathcal{O}_F \to \mathbf{Z}_p$ the projection on the v-component.

3.7.1. *The principal bundle.* Over Y_K we have a prime-to-p isogeny class of abelian schemes and therefore we have a canonical Barsotti–Tate group scheme $\mathcal{G}$. We let $\omega_{\mathcal{G}}$ be its conormal sheaf. The sheaf $\omega_{\mathcal{G}}$ carries an action of $\mathcal{O}_F$. We have a decomposition $\mathcal{O}_F \otimes_{\mathbf{Z}} \mathbf{Z}_p = \prod_{v|p} \mathbf{Z}_p$ and accordingly, the sheaf $\omega_{\mathcal{G}}$ decomposes as a product: $\omega_{\mathcal{G}} = \prod_{v|p} \omega_{\mathcal{G},v}$ where each $\omega_{\mathcal{G},v}$ is a locally free sheaf of rank 2 over Y_K.

3.7.2. *Weights for G and G_1.* By a dominant algebraic weight κ for G we mean a tuple $(k_v, l_v)_{v \in S_p}$ of integers such that $k_v \geq l_v$ for all $v \in S_p$. By a classical algebraic weight we mean a dominant algebraic weight which furthermore satisfies $l_v \geq 2$ for all $v \in S_p$. We will frequently write "weight" for "dominant algebraic weight" where no confusion can result (note though that we will later also consider p-adic weights). We associate a locally free sheaf ω^κ on Y_K to each weight κ by

$$\omega^\kappa = \prod_v \mathrm{Sym}^{k_v - l_v} \omega_{\mathcal{G},v} \otimes \det{}^{l_v} \omega_{\mathcal{G},v}.$$

By a weight κ for G_1 we mean a tuple $((k_v, l_v)_{v \in S_p}, w)$ of integers with the property that $k_v \geq l_v$ and $k_v - l_v \equiv w \pmod 2$ for each v; again, we say that κ is classical algebraic if $l_v \geq 2$ for all $v \in S_p$. In fact, we will insist that w is even, and we will shortly fix the choice $w = 2$. We claim that given w, there is a canonical descent datum on ω^κ for the map $Y_K \to Y_K^{G_1}$.

For clarity, we describe this descent datum on the level of the groupoid $\mathbf{Y}_K$. For all $x \in (\mathcal{O}_F)_{(p)}^{\times,+}$, we define an isomorphism

$$\omega^\kappa_{(A,\iota,x^{-1}\lambda,\eta,\eta_p)} = \omega^\kappa_{(A,\iota,\lambda,\eta,\eta_p)} \to \omega^\kappa_{(A,\iota,\lambda,\eta,\eta_p)}$$

by multiplication by $\prod_v v(x)^{(k_v+l_v-w)/2}$ (here the first identification is the tautological one, noting that the definition of ω^κ does not depend on the polarization).

To check that this defines a descent datum, we have to show that it respects the existing identifications from the action of $\mathcal{O}_F^{\times,+}(K^p)$. If $x \in \mathcal{O}_F^{\times,+}(K^p)$, then we may write $x = \epsilon^2$ for some $\epsilon \in \mathcal{O}_F^\times \cap K^p$, and we have an isomorphism $\epsilon : A \to A$ which induces an isomorphism in the groupoid $\mathbf{Y}_K$:

$$\epsilon : (A, \iota, \lambda, \eta, \eta_p) \to (A, \iota, \epsilon^{-2}\lambda, \eta, \eta_p)$$

and an isomorphism

$$\omega^\kappa_{(A,\iota,\epsilon^{-2}\lambda,\eta,\eta_p)} = \omega^\kappa_{(A,\iota,\lambda,\eta,\eta_p)} \xrightarrow{\epsilon^*} \omega_{(A,\iota,\lambda,\eta,\eta_p)}$$

which is multiplication by $\kappa(\epsilon)$ (again, the first equality is the tautological one, since ω^κ does not depend on the polarization). Now, $\kappa(\epsilon) = \prod_v v(\epsilon)^{k_v+l_v} = \prod_v v(\epsilon^2)^{(k_v+l_v-w)/2}\mathrm{N}_{F/\mathbf{Q}}(\epsilon)^w = \prod_v v(x)^{(k_v+l_v-w)/2}$ since $\mathrm{N}_{F/\mathbf{Q}}(\epsilon)^w = 1$ by our assumption that w is even, so this agrees with our the isomorphism defined above, as required.

This defines a descent datum for the étale map $Y_K \to Y_K^{G_1}$. This descent datum is effective. Indeed, after first identifying the sheaf ω^κ on various connected components of Y_K we are reduced to a finite étale descent for the group $\Delta(\Pi_0)$.

Although the descent datum depends on w, we will regard w as fixed (indeed, in the main arguments of the paper, we always take $w = 2$), so we omit it from the notation, and simply denote the resulting sheaf on $Y_K^{G_1}$ by ω^κ.

Remark 3.7.3. We assume in this remark that we work over $\mathbf{F}_p$ rather than $\mathbf{Z}_p$. We denote by $Y_{K,1}$ and $Y_{K,1}^{G_1}$ the fibres of Y_K and $Y_K^{G_1}$ over $\operatorname{Spec}\mathbf{F}_p$. Let $\kappa = (k_v, l_v)_{v|p}$ be a weight for G. We further assume that $k_v \equiv l_v \equiv 0 \mod (p-1)$. In this case, we claim that we can define a canonical descent datum for the sheaf ω^κ, from $Y_{K,1}$ to $Y_{K,1}^{G_1}$. This rests on the observation that the character $\mathcal{O}_F^\times \to \mathbf{F}_p^\times$ given by $\epsilon \mapsto \prod_{v|p}[v(\epsilon)^{k_v+l_v} \mod p]$ is trivial. Therefore we can define a descent datum for the action of $x \in (\mathcal{O}_F)_{(p)}^{\times,+}$, via the tautological isomorphism

$$\omega^\kappa_{(A,\iota,x^{-1}\lambda,\eta,\eta_p)} = \omega^\kappa_{(A,\iota,\lambda,\eta,\eta_p)}.$$

This remark will be applied to the various Hasse invariants we will construct later.

Finally we will need to consider the canonical extensions of these sheaves to toroidal compactifications. The conormal sheaf $\omega_{\mathcal{G}}/Y_K$ has a canonical extension to $X_{K,\Sigma}$ given by $e^*\Omega^1_{A/X_{K,\Sigma}}$, where A is the semi-abelian scheme of Theorem 3.5.1 (2) and e is its identity section. This gives an extension of the sheaves ω^κ to $X_{K,\Sigma}$ and an extension of the sheaves $\omega^{\kappa_{G_1}}$ to $X_{K,\Sigma}^{G_1}$. We will denote these extensions by the same symbol.

3.8. **Coherent cohomology and Hecke operators.**

3.8.1. *Basics.* Let $\kappa = (k_v, l_v)$ be a weight. We will study the cohomologies $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa)$ as well as their cuspidal variants $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa(-D))$ and $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa(-D))$.

Lemma 3.8.2. *The cohomologies* $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$, $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa)$, $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa(-D))$ *and* $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa(-D))$ *are independent of* Σ.

Proof. This is immediate from Theorem 3.5.1 (3). □

Because of this lemma, we often drop Σ from the notation. We now clarify the relationship between $\mathrm{R}\Gamma(X_K, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa)$.

Proposition 3.8.3. *The pull back maps*

$$\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa) \to \mathrm{R}\Gamma(X_K, \omega^\kappa)$$

and

$$\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa(-D)) \to \mathrm{R}\Gamma(X_K, \omega^\kappa(-D))$$

split in the derived category of $\mathbf{Z}_p$*-modules.*

Remark 3.8.4. It is often easier to work over X_K rather than $X_K^{G_1}$ because the former has a clear moduli interpretation. Proposition 3.8.3 tells us that we can easily transfer a good property of the cohomology over X_K to a property over $X_K^{G_1}$.

Proof of Proposition 3.8.3. Attached to the weight κ is a descent datum (see §3.7.2) which takes the form of an action of $(\mathcal{O}_F)_{(p)}^{\times,+}$ on the sheaf ω^κ over X_K. Namely, for all $\epsilon \in (\mathcal{O}_F)_{(p)}^{\times,+}$, there is an isomorphism $\epsilon : \epsilon^*\omega^\kappa \to \omega^\kappa$ satisfying the usual cocycle relation. This map induces a map on cohomology:

$$\epsilon : \mathrm{R}\Gamma(X_K, \omega^\kappa) \to \mathrm{R}\Gamma(X_K, \epsilon^*\omega^\kappa) \to \mathrm{R}\Gamma(X_K, \omega^\kappa)$$

and defines the group action.

Recall that there is a commutative diagram:

$$\begin{array}{ccc} X_K & \longrightarrow & X_K^{G_1} \\ \downarrow{\scriptstyle \Pi_0} & & \downarrow{\scriptstyle \Pi_0^{G_1}} \\ \mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1) & \xrightarrow{\pi} & (\mathcal{O}_F)_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1) \end{array}$$

Each Galois orbit $c \in [\mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)]/\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ determines a connected component of $\mathbf{Z}_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)$, and its fibre is a connected component $X_{K,c}$ of $X_{K,\Sigma}$ which is a proper scheme over $\operatorname{Spec}\mathbf{Z}_p$. Obviously $\mathrm{R}\Gamma(X_K, \omega^\kappa) = \prod_c \mathrm{R}\Gamma(X_{K,c}, \omega^\kappa)$ and for all $\epsilon \in (\mathcal{O}_F)_{(p)}^{\times,+}$, we have an isomorphism $\epsilon : \mathrm{R}\Gamma(X_{K,\epsilon\cdot c}, \omega^\kappa) \to \mathrm{R}\Gamma(X_{K,c}, \omega^\kappa)$.

The subgroup that fixes a component $X_{K,c}$ is denoted by $\mathcal{O}_F^{\times,+}(\Pi_0)$ and the action of this group on $X_{K,c}$ and $\mathrm{R}\Gamma(X_{K,c}, \omega^\kappa)$ actually factors through the finite group $\Delta(\Pi_0)$. Let $\pi(c)$ be the image of c in

$$[(\mathcal{O}_F)_{(p)}^{\times,+}\backslash(\mathbf{A}^{\infty,p}\otimes F)^\times/\nu(K^p)(1)]/\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}).$$

This determines a connected component $X_{K,\pi(c)}^{G_1}$ of $X_K^{G_1}$ and the map $X_{K,c} \to X_{K,\pi(c)}^{G_1}$ is a finite étale cover with group $\Delta(\Pi_0)$.

It follows from Lemma 3.8.5 below that $\mathrm{R}\Gamma(X_{K,\pi(c)}^{G_1}, \omega^\kappa)$ is split in $\mathrm{R}\Gamma(X_{K,c}, \omega^\kappa)$, and therefore the map

$$\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa) = \bigoplus_{\pi(c)} \mathrm{R}\Gamma(X_{K,\pi(c)}^{G_1}, \omega^\kappa) \to \prod_c \mathrm{R}\Gamma(X_{K,c}, \omega^\kappa) = \mathrm{R}\Gamma(X_K, \omega^\kappa)$$

is split. □

Lemma 3.8.5. *Let G be a finite group. Let $I_G \subset \mathbf{Z}[G]$ be the augmentation ideal. Let $f : T \to S$ be a finite étale morphism with Galois group G. Then $f_*\mathcal{O}_T = \mathcal{O}_S \oplus I_G \otimes_{\mathbf{Z}[G]} f_*\mathcal{O}_T$.*

Proof. There is an obvious map of coherent sheaves $\mathcal{O}_S \oplus I_G \otimes_{\mathbf{Z}[G]} f_*\mathcal{O}_T \to f_*\mathcal{O}_T$. The sheaf $f_*\mathcal{O}_T$ is a locally free sheaf (for the étale topology) of $\mathcal{O}_X[G]$-modules. Therefore, the above map is an isomorphism as this can be checked locally for the étale topology. □

3.8.6. *Abstract Hecke algebras.* Let $\mathcal{H} = \mathcal{C}_c^\infty(G_1(\mathbf{A}^\infty)//K, \mathbf{Z}_p)$ be the convolution algebra of locally constant, bi-K invariant, compactly supported functions on $G_1(\mathbf{A}^\infty)$ with coefficients in $\mathbf{Z}_p$. (The Haar measure is a product of local Haar measures, normalized by $\mathrm{vol}(K_t) = 1$ for all finite places t of F.) If S is a finite set of places of F, we let $\mathcal{H}^S$ be the subalgebra of $\mathcal{H}$ of functions whose restriction to $\mathrm{GSp}_4(F_s)$ is the characteristic function of K_s for all $s \in S$. For all finite places s, we let $\mathcal{H}_s$ be the local Hecke algebra $\mathcal{C}_c^\infty(\mathrm{GSp}_4(F_s)//K_s, \mathbf{Z}_p)$, so that $\mathcal{H} = \otimes_s' \mathcal{H}_s$.

3.8.7. *Cohomological correspondences — motivation.* We begin by giving some brief motivation for the way in which we define Hecke operators on coherent cohomology (following [Pil20]).

As usual, the geometric interpretation of Hecke operators is via correspondences

$$\begin{tikzcd} & C \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X & & Y \end{tikzcd}$$

(Giving an *integral* definition of the correspondence associated to a Hecke operator at a place dividing p is in general difficult. This question will be addressed later in the paper in some very special cases.)

Let $\mathcal{F}, \mathcal{G}$ be coherent sheaves on X, Y. We assume that we have a map of sheaves $p_2^*\mathcal{F} \to p_1^*\mathcal{G}$. When $\mathcal{F}$ and $\mathcal{G}$ are automorphic vector bundles (which will typically be the case for us), this map is provided by the differential of the universal isogeny over C.

One would like to use the correspondence to define a corresponding map on cohomology $\mathrm{R}\Gamma(X, \mathcal{F}) \to \mathrm{R}\Gamma(Y, \mathcal{G})$. This map could be defined by first taking the pull back via $p_2 : \mathrm{R}\Gamma(X, \mathcal{F}) \to \mathrm{R}\Gamma(C, p_2^*\mathcal{F})$, then using the map $p_2^*\mathcal{F} \to p_1^*\mathcal{G}$ to get to $\mathrm{R}\Gamma(C, p_1^*\mathcal{G})$, and finally applying some trace map to $\mathrm{R}\Gamma(Y, \mathcal{G})$. In other words, the action of the correspondence on cohomology should take the form of a map $T : R(p_1)_* p_2^*\mathcal{F} \to \mathcal{G}$. There are, however, at least two serious difficulties with making such a definition in our context.

The first obvious difficulty is the existence of the trace map, because in general one cannot assume that p_1 is finite flat. Nevertheless, in our cases the existence of the trace map will follow from the machinery of duality in coherent cohomology and the existence of certain fundamental classes, which can be constructed because the schemes C, X, Y will have reasonable geometric properties over the base.

The second difficulty (which already arises for modular forms for $\mathrm{GL}_2/\mathbf{Q}$) is that the action of the correspondences defining the Hecke operators at places dividing p is typically divisible by a positive power of p, so that one has to divide by this power in order to define the correct operator mod p. It is hard to check this divisibility at the level of the derived category.

The solution to this introduced in [Pil20] (which we also employ here) is as follows. By adjunction we can view T as a map $T : p_2^*\mathcal{F} \to p_1^!\mathcal{G}$, and in favourable circumstances $p_1^!\mathcal{G}$ will be a sheaf (and not merely a complex). Furthermore it will be sufficiently nice that we can check the condition that T is divisible by a power of p after restricting to the complement of a codimension 2 locus, and define our normalized Hecke operators.

3.8.8. *Duality for coherent complexes.* We let S be an affine Noetherian scheme. We say that a morphism $f : X \to Y$ of S-schemes is *embeddable* if there is a smooth S-scheme P such that f can be factored as a composite

$$X \xrightarrow{i} P \times_S Y \to Y$$

where i is finite and the second map is the natural projection. We say that f is *projectively embeddable* if p can be taken to be a projective space over S. In our applications of this material all of our maps will be obviously projectively embeddable (essentially because our Shimura varieties are quasi-projective), and we will not comment further upon this.

As usual we write $D_{qcoh}(\mathcal{O}_X)$ for the derived category of $\mathcal{O}_X$-modules with quasi-coherent cohomology sheaves, and $D^+_{qcoh}(\mathcal{O}_X)$ for the bounded-below version. Then if $f : X \to Y$ is an embeddable morphism of S-schemes, there is an exact functor of triangulated categories

$$f^! : D^+_{qcoh}(\mathcal{O}_Y) \to D^+_{qcoh}(\mathcal{O}_X).$$

If f is projectively embeddable, the functor $f^!$ is a right adjoint to Rf_* and there is a natural transformation $Rf_*f^! \implies \mathrm{Id}$ of endofunctors of $D^+_{qcoh}(\mathcal{O}_Y)$, which we refer to as the trace map.

If $X \to S$ is a local complete intersection then we write $K_{X/S}$ for the relative canonical sheaf, which may be defined as the determinant of the corresponding cotangent complex. The following is [Pil20, Cor. 4.1.3.1].

Lemma 3.8.9. *Let $f : X \to Y$ be an embeddable morphism between two embeddable S-schemes, such that $X \to S$, $Y \to S$ are both local complete intersections of pure relative dimension n. Then $f^!\mathcal{O}_Y = K_{X/S} \otimes_{\mathcal{O}_X} f^*K^{-1}_{Y/S}$ is an invertible sheaf.*

We will make repeated use of the following lemma.

Lemma 3.8.10. *Suppose that $f : X \to Y$ is an embeddable morphism of embeddable S-schemes, each of which is a local complete intersection of pure relative dimension n over S. Let h be a section of a line bundle $\mathcal{L}$ over Y, and suppose that neither h nor f^*h is a zero-divisor. Write $Y_{h=0}$ for the vanishing locus of h, and $X_{h=0}$ for the vanishing locus of f^*h.*

Then for any locally free sheaf $\mathcal{F}$ on Y, we have an equality of invertible sheaves

$$(f^!\mathcal{F})|_{X_{h=0}} = f^!(\mathcal{F}|_{Y_{h=0}}).$$

Proof. This follows from [Har66, Prop. III.8.8]. More precisely, note that $\mathcal{O}_{Y_{h=0}}$ is represented by the perfect complex of $\mathcal{O}_Y$-modules $\mathcal{L}^{-1} \xrightarrow{h} \mathcal{O}_Y$ (here we use that h is not a zero-divisor). In addition, by Lemma 3.8.9, $f^!\mathcal{F}$ is a sheaf, and it follows from the assumption that neither h nor f^*h is a zero-divisor that the derived tensor products in [Har66, Prop. III.8.8] are in our case given by the usual tensor product $\otimes$. $\square$

3.8.11. *Fundamental classes.* In two particular situations, we now construct a natural map

$$\Theta : \mathcal{O}_X = f^*\mathcal{O}_Y \to f^!\mathcal{O}_Y$$

which we call the *fundamental class*.

We firstly consider what we call the *lci situation*, which is the case that:

- X and Y are local complete intersections over S of the same relative dimension,
- X is normal, and
- there is an open $V \subset X$ which is smooth over S, whose complement is of codimension 2 in X, and an open $U \subset Y$ which is smooth and such that $f(V) \subset U$.

In this situation, $f^!\mathcal{O}_Y$ is an invertible sheaf by Lemma 3.8.9, so by the algebraic Hartogs' lemma, it is enough to specify the fundamental class over V (note that X is normal by assumption). Again by Lemma 3.8.9 we have $f^!\mathcal{O}_Y|_V = \det \Omega^1_{V/S} \otimes f^*(\det \Omega^1_{U/S})^{-1}$, so over V, we can define the fundamental class to be the determinant of the map

$$\mathrm{d}f : f^*\Omega^1_{U/S} \to \Omega^1_{V/S}.$$

The other case we consider is the *finite flat* situation, in which $f : X \to Y$ is a finite flat map, so that f_* is exact, and

$$f_*f^!\mathcal{O}_Y = \mathcal{H}om_{\mathcal{O}_Y}(f_*\mathcal{O}_X, \mathcal{O}_Y).$$

We have the usual trace morphism $\mathrm{tr}_f : f_*\mathcal{O}_X \to \mathcal{O}_Y$, and we define the fundamental class $f_*\mathcal{O}_X \to \mathcal{H}om_{\mathcal{O}_Y}(f_*\mathcal{O}_X, \mathcal{O}_Y)$ by $\Theta(1) = \mathrm{tr}_f$.

Note that if $X \to Y$ is a finite flat morphism and X, Y are both smooth over S, then the morphism $X \to Y$ is automatically a local complete intersection. The following compatibility between these definitions is [Pil20, Lem. 4.2.3.1].

Lemma 3.8.12. *Suppose that $X \to Y$ is finite flat, and that X, Y are both smooth over S. Then*

$$\mathbf{L}_{X/Y} \xrightarrow{\sim} [\Omega^1_{Y/S} \otimes_{\mathcal{O}_Y} \mathcal{O}_X \xrightarrow{df} \Omega^1_{X/S}],$$

and the determinant $\det(df) \in \omega_{X/Y} = f^!\mathcal{O}_Y$ *is the trace map* tr_f.

3.8.13. *Base change for open immersions.* Consider a Cartesian diagram

$$\begin{tikzcd} X' \arrow[r, "j"] \arrow[d, "f'"] & X \arrow[d, "f"] \\ Y' \arrow[r, "i"] & Y \end{tikzcd}$$

If i is an open immersion, and f is in either of the finite flat or lci situations, then so is f'. Since $i^! = i^*$ and $j^! = j^*$, we have $j^*f^! = (f')^!i^*$, and if f has fundamental class Θ, then $j^*\Theta$ is the fundamental class of f'.

3.8.14. *Fundamental classes and divisors.* We now briefly recall the results of [Pil20, §4.2.4], which show that the correspondences we define below are suitably well behaved on the boundaries of our compactified Shimura varieties.

Let $D_X \hookrightarrow X$, $D_Y \hookrightarrow Y$ be two effective reduced Cartier divisors with respect to S, with the properties that $f : X \to Y$ restricts to a map $f|_{D_X} : D_X \to D_Y$, and the induced map $D_X \to f^{-1}(D_Y)$ is an isomorphism of topological spaces. Write X^{sm}, Y^{sm} for the smooth loci of X, Y. The following is [Pil20, Lem. 4.2.4.1].

Lemma 3.8.15. *Suppose either that we are in the finite flat situation; or that we are in the lci situation and that furthermore $D_X \cap X^{sm}$ and $D_Y \cap Y^{sm}$ are normal crossings divisors.*

Then the fundamental class $\Theta : \mathcal{O}_X \to f^!\mathcal{O}_Y$ restricts to a morphism $\mathcal{O}_X(-D_X) \to f^!\mathcal{O}_Y(-D_Y)$.

3.8.16. *Traces and restriction.* In this paper we will have to study how Hecke operators behave with respect to restriction to subschemes of the Shimura variety. This section contains some preliminary material. Consider the following setup:

- $f : X \to Y$ a finite flat map between smooth varieties over a field k.
- $D \subset Y$ is a smooth Cartier divisor.
- $f^{-1}(D) = nD'$ for $D' \subset X$ a smooth Cartier divisor.

In this setting we have the following:

- Trace maps on canonical bundles
$$f_*K_X \to K_Y$$
and
$$f_*K_{D'} \to K_D.$$
- Adjunction isomorphisms
$$K_D \simeq K_Y(D)|_D$$
and
$$K_{D'} \simeq K_X(D')|_{D'}.$$

If $\mathcal{L}$ is a line bundle on Y, we can use the projection formula to get a map: $f_*(K_X \otimes_{\mathcal{O}_X} f^*\mathcal{L}) \to K_Y \otimes_{\mathcal{O}_Y} \mathcal{L}$. We call such a map a twisted trace map. We use a similar terminology over D. The goal of this section is to prove the following compatibility between them.

Proposition 3.8.17. *There is a commutative diagram*

$$\begin{tikzcd}
f_*(K_X(-(n-1)D')) \arrow[r] \arrow[d] & K_Y \arrow[d] \\
f_*(K_{D'} \otimes \mathcal{O}_X(-nD')|_{D'}) \arrow[r] & K_D \otimes \mathcal{O}_Y(-D)|_D
\end{tikzcd}$$

Here the vertical maps are restriction followed by adjunction, the top horizontal map comes from the inclusion of $K_X(-(n-1)D')$ in K_X followed by the trace, and the bottom horizontal arrow is the twisted trace for $f : D' \to D$ and the line bundle $\mathcal{O}_Y(-D)|_D$ (note that $f^*\mathcal{O}_Y(-D)|_D = \mathcal{O}_X(-nD)|_{D'}$).

Proof. We write $\mathcal{I} = \mathcal{O}_Y(-D)$ for the ideal sheaf of D and $\mathcal{I}' = \mathcal{O}_X(-D')$ for the ideal sheaf of D'. First consider the following commutative diagram:

$$\begin{tikzcd}
f_*\mathcal{I}'^{n-1}\,\mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X, \mathcal{O}_Y) \arrow[r, hook] \arrow[d, dotted] & f_*\,\mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X, \mathcal{O}_Y) \arrow[r] \arrow[d] & \mathcal{O}_Y \arrow[d] \\
f_*\,\mathcal{H}om_{\mathcal{O}_Y/\mathcal{I}}(\mathcal{O}_X/\mathcal{I}', \mathcal{O}_Y/\mathcal{I}) \arrow[r, hook] & f_*\,\mathcal{H}om_{\mathcal{O}_Y/\mathcal{I}}(\mathcal{O}_X/\mathcal{I}\mathcal{O}_X, \mathcal{O}_Y/\mathcal{I}) \arrow[r] & \mathcal{O}_Y/\mathcal{I}
\end{tikzcd}$$

where $\mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X, \mathcal{O}_Y)$ is sheaf of $\mathcal{O}_Y$-homomorphisms from $f_*\mathcal{O}_X$ to $\mathcal{O}_Y$, which we view as a coherent sheaf of $\mathcal{O}_X$-modules. By definition $\mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X, \mathcal{O}_Y) = f^!\mathcal{O}_Y$.

Consider first the square on the right: the horizontal maps are evaluation at 1, while the vertical maps are given by reduction modulo $\mathcal{I}$, and it is clear that this square commutes.

Now we consider the left hand square: the horizontal maps are the obvious inclusions so we must explain why the dotted arrow exists. But a local section s of $\mathcal{I}'^{n-1}\,\mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X,\mathcal{O}_Y)$ will send $\mathcal{I}'$ into $\mathcal{I}$ (using that $\mathcal{I}'^n = \mathcal{I}\mathcal{O}_X$) and hence the reduction of s mod $\mathcal{I}$ factors through $\mathcal{O}_X/\mathcal{I}'$.

Finally we note that the square in the statement of the proposition tensored with K_Y^{-1} may be identified with the outer rectangle of this diagram because we have $K_X \otimes K_Y^{-1} \simeq f^!\mathcal{O}_Y = \mathcal{H}om_{\mathcal{O}_Y}(\mathcal{O}_X,\mathcal{O}_Y)$. □

3.9. **Cohomological correspondences — definitions.** Let S be a Noetherian scheme. Let X, Y be two S-schemes.

Definition 3.9.1. A *correspondence* C over X and Y is a diagram of S-morphisms:

$$\begin{tikzcd} & C \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X & & Y \end{tikzcd}$$

where X, Y, C have the same pure relative dimension over S and the morphisms p_1 and p_2 are projectively embeddable.

Let $\mathcal{F}$ be a coherent sheaf over X and $\mathcal{G}$ a coherent sheaf over Y.

Definition 3.9.2. A *cohomological correspondence* from $\mathcal{F}$ to $\mathcal{G}$ is the data of a correspondence C over X and Y and a map $T : \mathrm{R}(p_1)_* p_2^*\mathcal{F} \to \mathcal{G}$.

The map T can be seen, by adjunction, as a map $p_2^*\mathcal{F} \to p_1^!\mathcal{G}$. It gives rise to a map still denoted by T on cohomology:

$$\mathrm{R}\Gamma(X,\mathcal{F}) \xrightarrow{p_2^*} \mathrm{R}\Gamma(C, p_2^*\mathcal{F}) = \mathrm{R}\Gamma(Y, \mathrm{R}(p_1)_* p_2^*\mathcal{F}) \xrightarrow{T} \mathrm{R}\Gamma(Y,\mathcal{G}).$$

3.9.3. *Hecke action away from p.* Let $K = K^pK_p$ be a reasonable compact open subgroup of $G_1(\mathbf{A}_f)$. Let $\mathcal{H}^p = \mathcal{C}_c^\infty(G_1(\mathbf{A}^{p,\infty})//K^p, \mathbf{Z}_p)$ be the Hecke algebra away from p.

We claim that there is an action of $\mathcal{H}^p$ on $\mathrm{R}\Gamma(X_{K,\Sigma},\omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1},\omega^\kappa)$. To this end, let $g \in G_1(\mathbf{A}^{\infty,p})$. We will define an endomorphism of $\mathrm{R}\Gamma(X_K,\omega^\kappa)$ which corresponds to the action of the double class $[K^pgK^p]$.

We define (for suitable choices of cone decompositions omitted from the notation) a correspondence:

$$\begin{tikzcd} & X_{K\cap gKg^{-1}} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_K & & X_K \end{tikzcd}$$

where p_1 is the map induced from the inclusion $K\cap gKg^{-1} \subset K$ and the functoriality of §3.6.1.

The map p_2 is the composite of the map $[g] : X_{K\cap gKg^{-1}} \to X_{K\cap g^{-1}Kg}$ (see 3.6.2) and the natural map $X_{K\cap g^{-1}Kg} \to X_K$ deduced from the inclusion $K\cap gKg^{-1} \subset K$ and functoriality of §3.6.1.

We have a canonical isomorphism $p_2^*\omega^\kappa \xrightarrow{\sim} p_1^*\omega^\kappa$, because the construction of the sheaf ω^κ depends only on the p-divisible group. Moreover, because X_K and $X_{K\cap gKg^{-1}}$ are lci and smooth outside codimension 2 (for a cofinal subset of the set of all polyhedral cone decompositions), there is a fundamental class $p_1^*\mathcal{O}_{X_K} \to p_1^!\mathcal{O}_{X_K}$, extending the trace for the finite étale map p_1 on the interior, which we can tensor with $p_1^*\omega^\kappa$ to obtain a map $p_1^*\omega^\kappa \to p_1^!\omega^\kappa = p_1^!\mathcal{O}_{X_K} \otimes p_1^*\omega^\kappa$.

Composing the maps $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ and $p_1^*\omega^\kappa \to p_1^!\omega^\kappa$ we obtain a cohomological correspondence $\Theta_g : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$ which induces the operator $[KgK]$ on cohomology:

$$\mathrm{R}\Gamma(X_K, \omega^\kappa) \to \mathrm{R}\Gamma(X_{K\cap gKg^{-1}}, p_2^*\omega^\kappa) \xrightarrow{\Theta_g} \mathrm{R}\Gamma(X_{K\cap gKg^{-1}}, p_1^!\omega^\kappa) \xrightarrow{\mathrm{Tr}} \mathrm{R}\Gamma(X_K, \omega^\kappa)$$

where the last map is induced by the adjunction $\mathrm{Tr} : \mathrm{R}(p_1)_* p_1^!\omega^\kappa \to \omega^\kappa$.

We have a similar definition on cuspidal cohomology. Moreover, all these definitions commute with the action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ and therefore we also get an action on the cohomology $\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_K^{G_1}, \omega^\kappa(-D))$.

The characteristic functions of the double classes $[K^p g K^p]$ generate $\mathcal{H}^p$ as a $\mathbf{Z}_p$-module. In Proposition 3.9.15 below we prove that when $K_p = \prod_{v|p} \mathrm{GSp}_4(\mathbf{Z}_p)$ is spherical, the actions we just defined of the $[K^p g K^p]$ are compatible with products in $\mathcal{H}^p$ (the composite action of $[K^p g_1 K^p]$ and $[K^p g_2 K^p]$ is equal to the action of $[K^p g_1 K^p][K^p g_2 K^p]$ decomposed into sum of elementary double classes) so that we get an action of the Hecke algebra $\mathcal{H}^p$.

The difficulties come from the boundary. Away from the boundary, all the correspondences are finite étale and one can follow the discussion of [FC90, Chap. VII, §3], to show the compatibility. Following that reference, it should be possible to show in a similar fashion that the action of the double class is compatible with product in the Hecke algebra on the compactified Shimura variety, but giving all the details would involve a delicate study of the composition of the correspondences at the boundary. We instead give a different ad hoc proof by exhibiting special complexes computing the cohomology. These complexes are Cousin complexes associated with the Ekhedal–Oort stratification on the Shimura variety. The action of all double classes $[K^p g K^p]$ on the cohomology is given by a canonical action on the complex. Moreover, each term of the complex is the global sections of a certain sheaf and the restriction of the sections of this sheaf to the interior of the Shimura variety is an embedding. We are therefore able to prove that the action of the double classes is compatible with products in the Hecke algebra because we know this holds on the non-compact Shimura variety.

Remark 3.9.4. Over $\mathbf{Q}_p$, the property that the action of the double class is compatible with product in the Hecke algebra follows from [Har90b, Prop. 2.6]. The strategy of that paper is to define an action of the group $\mathrm{G}_1(\mathbf{A}_f^p)$ after passing to the limit over the level K^p and then deduce an action of the Hecke algebra at a finite level, but this strategy requires more work over $\mathbf{Z}_p$ because at some points one needs to control the cohomology of finite groups (which vanishes in characteristic zero). Nevertheless, this is enough to prove that the Hecke algebra $\mathcal{H}^p$ acts on the torsion free part of the cohomology (which embeds in the cohomology with $\mathbf{Q}_p$-coefficients).

3.9.5. *Cousin complexes.* Our main reference for this section is [Kem78]. Let X be a topological space. Let $Sh_X(Ab)$ be the category of abelian sheaves on X. For

a subset $Z \subseteq X$ and abelian sheaf $\mathcal{F}$ we denote by $\Gamma_Z(\mathcal{F})$ the subsheaf of $\mathcal{F}$ of sections supported on Z. Let $Z : Z_0 = X \supseteq Z_1 \supseteq \cdots \supseteq Z_n \cdots$ be a decreasing sequence of closed subsets of X (called a filtration). For any abelian sheaf $\mathcal{F}$ on X, one can build the Cousin complex of $\mathcal{F}$ with respect to the filtration Z, denoted by $\mathrm{Cous}_Z(\mathcal{F})$ [Kem78, p. 357].

The Cousin complex $\mathrm{Cous}_Z(\mathcal{F})$ is a complex of abelian sheaves in positive degree. The object in degree i is $\mathcal{H}^i_{Z_i/Z_{i+1}}(\mathcal{F})$, where $\mathcal{H}^k_{Z_i/Z_{i+1}}(\cdot)$ is (by [Kem78, Lem. 7.3]) the k-th derived functor of the functor:

$$\begin{array}{rcl} \mathrm{Sh}_X(Ab) & \to & \mathrm{Sh}_X(Ab) \\ \mathcal{G} & \mapsto & [U \mapsto \Gamma_{Z_i \setminus Z_{i+1}}(U \setminus Z_{i+1}, \mathcal{G})] \end{array}$$

The differential $\mathcal{H}^i_{Z_i/Z_{i+1}}(\mathcal{F}) \to \mathcal{H}^{i+1}_{Z_{i+1}/Z_{i+2}}(\mathcal{F})$ is induced by a certain boundary map. The Cousin complex has an augmentation $\mathcal{F} \to \mathrm{Cous}(\mathcal{F})$.

We now specialize the discussion: X is a Noetherian scheme and $\mathcal{F}$ is a quasi-coherent sheaf. Then $\mathrm{Cous}(\mathcal{F})$ is a complex of quasi-coherent sheaves.

We have the following theorem:

Theorem 3.9.6. *Let X be a Noetherian scheme with a filtration Z by closed subschemes that satisfies:*

(1) $\mathrm{codim}_X(Z_i) \geq i$.
(2) The morphism $Z_i \setminus Z_{i+1} \to X$ is affine for all i.

Let $\mathcal{F}$ be a maximal Cohen–Macaulay coherent sheaf on X. Then $\mathrm{Cous}_Z(\mathcal{F})$ *is quasi-isomorphic to $\mathcal{F}$.*

Proof. This follows from [Kem78, Thm. 10.9] (by definition a sheaf $\mathcal{F}$ is locally Cohen–Macaulay with respect to a filtration Z if $\mathrm{Cous}_Z(\mathcal{F})$ is a resolution of $\mathcal{F}$, see [Kem78, p. 358]). □

Remark 3.9.7. If we further assume that each $Z_i \setminus Z_{i+1}$ is affine, then $\mathrm{Cous}_Z(\mathcal{F})$ is a complex of acyclic sheaves by [Kem78, Thm. 9.6].

One can sometimes compute the complex $\mathrm{Cous}_Z(\mathcal{F})$ more explicitly. Write $U_{i+1} = X \setminus Z_{i+1}$, and write $j_{i+1} : U_{i+1} \hookrightarrow X$ for the inclusion. Under the assumption that $Z_i \setminus Z_{i+1} \to X$ is affine, we have by [Kem78, Lem. 8.5(e)] (note that the spectral sequence there degenerates by [Kem78, Thm. 9.6(c)], as in the proof of [Kem78, Thm. 9.5]):

$$\mathcal{H}^k_{Z_i/Z_{i+1}}(\mathcal{F}) = (j_{i+1})_* \mathrm{R}^k \Gamma_{Z_i \setminus Z_{i+1}}(\mathcal{F}|_{U_{i+1}}). \tag{3.9.8}$$

In general, for a Noetherian scheme X and a closed subset Z defined by an ideal sheaf $\mathcal{I}$, we have

$$\mathrm{R}^i \Gamma_Z(\mathcal{F}) = \varinjlim_n \mathcal{E}xt^i(\mathcal{O}_X/\mathcal{I}^n, \mathcal{F})$$

and these $\mathcal{E}xt$ sheaves can be computed by taking projective resolutions of $\mathcal{O}_X/\mathcal{I}^n$. We also remark that in the previous limit, we can replace the ideals $\mathcal{I}^n$ by any other decreasing sequence of ideals $\{J_n\}$ with the property that for all n, there is k and k' such that $J_{k'} \subset I^n \subset J_k$.

Example 3.9.9. We are going to compute these $\mathcal{E}xt$ sheaves in a special case. Assume that we have effective Cartier divisors $\mathcal{O}_X \xrightarrow{s_t} \mathcal{L}_t$ for $1 \leq t \leq i$ and assume that they intersect properly, by which we mean that for all n, the "twisted" Koszul complex:

$$\mathrm{Kos}(s_1^n, \cdots, s_i^n) : 0 \to \bigotimes_t \mathcal{L}_t^{-n} \to \oplus_{t'} \bigotimes_{t \neq t'} \mathcal{L}_t^{-n} \to \cdots \to \oplus_t \mathcal{L}_t^{-n} \to \mathcal{O}_X \to 0$$

is a projective resolution of $\mathcal{O}_X/(\mathcal{L}_1^{-n}, \cdots, \mathcal{L}_t^{-n})$.

We let $Z = V(\mathcal{L}_1^{-1}, \cdots, \mathcal{L}_i^{-1})$ and let $\mathcal{F}$ be a locally free coherent sheaf. We find that: $\mathcal{E}xt^j(\mathcal{O}_X/(\mathcal{L}_1^{-n}, \cdots, \mathcal{L}_i^{-n}), \mathcal{F}) = 0$ unless $j = i$, and

$$\mathcal{E}xt^i(\mathcal{O}_X/(\mathcal{L}_1^{-n}, \cdots, \mathcal{L}_i^{-n}), \mathcal{F}) = \mathrm{Coker}(\oplus_t s_t^n : \bigotimes_{t' \neq t} \mathcal{L}_{t'}^n \otimes \mathcal{F} \to (\bigotimes_t \mathcal{L}_t^n) \otimes \mathcal{F}).$$

Taking the direct limit over n gives $\mathrm{R}^i\Gamma_Z(\mathcal{F})$.

3.9.10. *The Cousin complex of the Ekedahl–Oort stratification.* We now assume that $K_p = \prod_v \mathrm{GSp}_4(\mathbf{Z}_p)$, and let $X = X_{K,\Sigma}$ and denote by X^* the minimal compactification. We have a morphism $f : X \to X^*$. We fix an integer n and work over $X_n = X \times \mathrm{Spec}\, \mathbf{Z}/p^n\mathbf{Z}$ and $X_n^* = X^* \times \mathrm{Spec}\, \mathbf{Z}/p^n\mathbf{Z}$, and let Y_n denote the interior of X_n. We consider the filtration Z^* on X_n^* given by taking Z_i^* to be the closure of all Ekedahl–Oort strata of codimension i. Here are some known facts (see [Box15, Thm. 6.2.3]):

(1) Z^* is a filtration.
(2) $Z_i^* \setminus Z_{i+1}^*$ is affine.
(3) $Z_i^* \setminus Z_{i+1}^*$ is a set-theoretic local complete intersection in $U_{i+1}^* = X_n^* \setminus Z_{i+1}^*$.

We now consider the pull-back Z of Z^* on X_n. We deduce that:

(1) Z is a filtration.
(2) $Z_i \setminus Z_{i+1} \hookrightarrow X_n$ is affine.
(3) $Z_i \setminus Z_{i+1}$ is a set-theoretic local complete intersection in $U_{i+1} = X_n \setminus Z_{i+1}$.

Proposition 3.9.11. *The cohomology* $\mathrm{R}\Gamma(X_n, \omega^\kappa(-D))$ *is computed by*

$$\Gamma(X_n, \mathrm{Cous}_Z(\omega^\kappa(-D))).$$

Proof. It follows from Theorem 3.9.6 that $\omega^\kappa(-D) \to \mathrm{Cous}_Z(\omega^\kappa(-D))$ is a quasi-isomorphism. It suffices to prove that $\mathrm{Cous}_Z(\omega^\kappa(-D))$ is a complex of acyclic sheaves. By (3.9.8), the sheaf in degree i is equal to $(j_{i+1})_*\mathrm{R}^i\Gamma_{Z_i \setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}})$. This sheaf is supported on $Z_i \setminus Z_{i+1}$. We claim that $Rf_*(j_{i+1})_*\mathrm{R}^i\Gamma_{Z_i \setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}})$ is concentrated in degree 0. Since $f_*(j_{i+1})_*\mathrm{R}^i\Gamma_{Z_i \setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}})$ is an acyclic sheaf because it is supported on $Z_i^* \setminus Z_{i+1}^*$, the proposition will follow from our claim.

Let us prove the claim. By construction, $Z_i \setminus Z_{i+1}$ is a finite disjoint union of Ekedahl–Oort strata and $(j_{i+1})_*\mathrm{R}^i\Gamma_{Z_i \setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}})$ is a finite direct sum indexed by these Ekedahl–Oort strata. Let E be an Ekedahl–Oort stratum appearing in $Z_i \setminus Z_{i+1}$. It can be written as the intersection of i Cartier divisors in U_{i+1}, using the theory of generalized Hasse invariants (one can also assume that these Cartier divisors are pulled back from X_n^*; note that a sufficiently large power of each generalized Hasse invariant can be lifted to X_n^*). Let us denote these Cartier divisors by $\mathcal{O}_X \xrightarrow{s_t} \mathcal{I}_t$. It follows from Example 3.9.9 that the direct summand of the sheaf $\mathrm{R}^i\Gamma_{Z_i \setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}})$ corresponding to E is the inductive limit of the sheaves:

$$\mathcal{H}^i\mathcal{H}om(\mathrm{Kos}(s_1^n, \cdots, s_i^n), \omega^\kappa(-D)).$$

The complex $\mathcal{H}om(\mathrm{Kos}(s_1^n, \cdots, s_i^n), \omega^\kappa(-D))$ is a complex of sheaves acyclic relatively to the minimal compactification, and concentrated in degree i. $\square$

Lemma 3.9.12. *There is an injection of complexes*

$$\Gamma(X_n, \mathrm{Cous}_Z(\omega^\kappa(-D))) \hookrightarrow \Gamma(Y_n, \mathrm{Cous}_Z(\omega^\kappa(-D))).$$

Proof. This follows directly from the description of the objects of the complex $\mathrm{Cous}_Z(\omega^\kappa(-D))$ given in the course of the preceding proof. □

It remains to prove that our Hecke operators act on $\Gamma(X_n, \mathrm{Cous}_Z(\omega^\kappa(-D)))$ and $\Gamma(Y_n, \mathrm{Cous}_Z(\omega^\kappa(-D)))$. Let $g \in G(\mathbf{A}_f^p)$. We consider the correspondence:

$$\begin{tikzcd} & X_{K\cap gKg^{-1}} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_K & & X_K \end{tikzcd}$$

and more precisely its reduction modulo p^n. We have a cohomological correspondence $p_2^*\omega^\kappa(-D) \to p_1^!\omega^\kappa(-D)$, as defined in §3.9.3.

Lemma 3.9.13. *This cohomological correspondence induces a cohomological correspondence of complexes compatible with the augmentation:*

$$p_2^*\mathrm{Cous}_Z(\omega^\kappa(-D)) \to p_1^!\mathrm{Cous}_Z(\omega^\kappa(-D))$$

Remark 3.9.14. In the above correspondence, the functors p_2^* and $p_1^!$ are applied to each object of the complex. Moreover, for each object $\mathrm{Cous}_Z(\omega^\kappa(-D))^i$ of $\mathrm{Cous}_Z(\omega^\kappa(-D))$, $p_1^!\mathrm{Cous}_Z(\omega^\kappa(-D))^i$ is a sheaf (i.e it is concentrated in degree 0).

Proof of Lemma 3.9.13. For each index i, we have

$$\mathrm{Cous}_Z(\omega^\kappa(-D))^i = (j_{i+1})_*\mathrm{R}^i\Gamma_{Z_i\setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}}).$$

We choose (by considering powers of generalized Hasse invariants) an increasing sequence $(Z_i \setminus Z_{i+1})_k$ of subschemes U_{i+1} with support $Z_i \setminus Z_{i+1}$, which are local complete intersections, and are cofinal among all subschemes of U_{i+1} with support $Z_i \setminus Z_{i+1}$.

We have $\mathrm{R}^i\Gamma_{Z_i\setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}}) = \varinjlim_k \mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}})$. Also recall that $\mathcal{E}xt(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}}) = \mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}})[-i]$.

We have

$$p_1^!(\mathcal{E}xt(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}}) = \mathcal{E}xt(p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, p_1^!\omega^\kappa(-D)|_{U_{i+1}})$$

by [Har66, Prop. III.8.8]. One checks that $p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k} = {}^{\mathbf{L}}p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}$ because the pull back of a local regular sequence defining $(Z_i\setminus Z_{i+1})_k$ is again a local regular sequence; we will not comment on the vanishing of higher pullbacks in the rest of this argument. We deduce that

$$p_1^!(\mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}}) = \mathcal{E}xt^i(p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, p_1^!\omega^\kappa(-D)|_{U_{i+1}}).$$

On the other hand there is by adjunction a map:

$$p_2^*\mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}}) \to \mathcal{E}xt^i(p_2^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, p_2^*\omega^\kappa(-D)|_{U_{i+1}}).$$

Since the Ekedahl–Oort stratification is invariant under prime to p isogenies, we deduce that $p_2^{-1}(Z_i \setminus Z_{i+1}) = p_1^{-1}(Z_i \setminus Z_{i+1})$. Therefore, for each k, for all large enough t, there is a natural map $p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_t} \to p_2^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}$.

We therefore get a map:

$$p_2^*\mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, \omega^\kappa(-D)|_{U_{i+1}}) \to \mathcal{E}xt^i(p_2^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_k}, p_2^*\omega^\kappa(-D)|_{U_{i+1}})$$

$$\to \mathcal{E}xt^i(p_1^*\mathcal{O}_{(Z_i\setminus Z_{i+1})_t}, p_1^!\omega^\kappa(-D)|_{U_{i+1}}) = p_1^!(\mathcal{E}xt^i(\mathcal{O}_{(Z_i\setminus Z_{i+1})_t}, \omega^\kappa(-D)|_{U_{i+1}}).$$

Passing to the inductive limit over k and t yields the cohomological correspondence:

$$p_2^*\mathrm{R}^i\Gamma_{Z_i\setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}}) \to p_1^!\mathrm{R}^i\Gamma_{Z_i\setminus Z_{i+1}}(\omega^\kappa(-D)|_{U_{i+1}}).$$

Moreover this construction is canonical and is compatible with all differentials in the Cousin complex and with the augmentation. □

Proposition 3.9.15. *The Hecke algebra $\mathcal{H}^p$ acts on $\mathrm{R}\Gamma(X^{G_1}, \omega^\kappa(-D))$ and also on $\mathrm{R}\Gamma(X^{G_1}, \omega^\kappa)$.*

Proof. By Serre duality, it suffices to treat the case of $\mathrm{R}\Gamma(X^{G_1}, \omega^\kappa(-D))$. The cohomology $\mathrm{R}\Gamma(X, \omega^\kappa(-D))$ is represented by

$$\varprojlim_n \Gamma(X_n, \mathrm{Cous}_Z(\omega^\kappa(-D)))$$

and this complex injects into

$$\varprojlim_n \Gamma(Y_n, \mathrm{Cous}_Z(\omega^\kappa(-D))).$$

The double class $\mathbf{1}_{K^pgK^p}$ acts everywhere. We can restrict to the "G_1" direct factor (the Ekedahl–Oort stratification is preserved by the action of $(\mathcal{O}_F)^{\times,+}_{(p)}$) and the compatibility with the product in the Hecke algebra is known on

$$\varprojlim_n \Gamma(Y_n^{G_1}, \mathrm{Cous}_Z(\omega^\kappa(-D))).$$

Therefore it holds everywhere. □

Remark 3.9.16. It follows by an identical argument to the proof of Proposition 3.9.15 that if K_1^p, K_2^p, K_3^p are three choices of tame level, and Σ_1, Σ_2, Σ_3 are suitable choices of polyhedral cone decompositions, then the composite of the Hecke operators

$$[K_1^pg_1K_2^p] : \mathrm{R}\Gamma(X^{G_1}_{K_2^pK_p,\Sigma_2}, \omega^\kappa) \to \mathrm{R}\Gamma(X^{G_1}_{K_1^pK_p,\Sigma_1}, \omega^\kappa)$$

and

$$[K_2^pg_2K_3^p] : \mathrm{R}\Gamma(X^{G_1}_{K_3^pK_p,\Sigma_3}, \omega^\kappa) \to \mathrm{R}\Gamma(X^{G_1}_{K_2^pK_p,\Sigma_2}, \omega^\kappa)$$

is the Hecke operator

$$[K_1^pg_1K_2^p][K_2^pg_2K_3^p] : \mathrm{R}\Gamma(X^{G_1}_{K_3^pK_p,\Sigma_3}, \omega^\kappa) \to \mathrm{R}\Gamma(X^{G_1}_{K_1^pK_p,\Sigma_1}, \omega^\kappa).$$

3.9.17. *Hecke operators at p: Siegel type operator.* We assume here that $K = K^pK_p$ and $K_p = G_1(\mathbf{Z}_p)$. Let us fix a place w above p. We are going to define an action of a Hecke operator $T_{w,1}$ on $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$, $\mathrm{R}\Gamma(X^{G_1}_{K,\Sigma}, \omega^\kappa)$, and their cuspidal versions. The action on $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa(-D))$ is not canonical, and depends on the choice of x_w made in §3.6.3, but the action on $\mathrm{R}\Gamma(X^{G_1}_{K,\Sigma}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X^{G_1}_{K,\Sigma}, \omega^\kappa(-D))$ is canonical.

Set $K' = K^pK'_p$ where $K'_p = \prod_{v\neq w} \mathrm{GSp}_4(\mathcal{O}_{F_v}) \times \mathrm{Si}(w)$. In §3.6.7 we defined maps $p_1, p_2 : X_{K'} \to X_K$ giving a Hecke correspondence:

$$\begin{tikzcd} & X_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_K & & X_K \end{tikzcd}$$

The key geometric properties of this correspondence are (see Proposition 3.4.1):

(1) $X_{K'}$ and X_K are relative complete intersections over $\operatorname{Spec}\mathbf{Z}_p$, and are pure of the same dimension,
(2) X_K is smooth over $\operatorname{Spec}\mathbf{Z}_p$,
(3) $X_{K'}$ is smooth over $\operatorname{Spec}\mathbf{Z}_p$ up to codimension 2 and normal.

In particular, we are in the lci situation in the sense of §3.8.11, so we have an invertible dualizing sheaf $p_1^!\mathcal{O}_{X_K}$ and a fundamental class $p_1^*\mathcal{O}_{X_K} \to p_1^!\mathcal{O}_{X_K}$. Moreover, for all weights $\kappa = (k_v, l_v)$ with $l_v \geq 0$, we have a natural map $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ provided by the differential of the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ on $X_{K'}$.

Composing these maps, we obtain a cohomological correspondence $\Theta : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$.

Lemma 3.9.18. *When $l_w \geq 2$, this map is divisible by p^3.*

Proof. We need to prove that Θ factors through $p^3 p_1^!\omega^\kappa$. As $X_{K'}$ is normal and the source and target are locally free sheaves, it is enough to establish this factorization in codimension one. As this factorization is furthermore trivial over the generic fibre of $X_{K'}$, it is enough to prove it over the completed local rings of the generic points of the special fibre of $X_{K'}$.

There are three types of generic points in the special fibre classified by the multiplicative rank $r = 0, 1$ or 2 of the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$. In each case one calculates separately the p-divisibility of the map $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ and of the fundamental class as in the proof of [Pil20, Lem. 7.1.1]. One finds that the fundamental class $p_1^*\mathcal{O}_{X_K} \to p_1^!\mathcal{O}_{X_K}$ is divisible by p^3 when $r = 0$, p when $r = 1$, and p^0 when $r = 2$, and the map $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ is divisible by p^0 when $r = 0$, p^{l_w} when $r = 1$, and $p^{k_w + l_w}$ when $r = 2$. The result follows as $3, l_w + 1, k_w + l_w \geq 3$. □

We can thus consider the normalized cohomological correspondence $T_{w,1} : p^{-3}\Theta : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$, and we obtain a Hecke operator:

$$T_{w,1} : \mathrm{R}\Gamma(X_K, \omega^\kappa) \to \mathrm{R}\Gamma(X_{K'}, p_2^*\omega^\kappa) \xrightarrow{p^{-3}\Theta} \mathrm{R}\Gamma(X_{K'}, p_1^!\omega^\kappa) \xrightarrow{\mathrm{Tr}} \mathrm{R}\Gamma(X_K, \omega^\kappa).$$

A similar definition applies to cuspidal cohomology and works over $X_K^{G_1}$.

Remark 3.9.19. One readily checks that the Hecke correspondence used to define $T_{w,1}$ corresponds to the double coset $[\mathrm{GSp}_4(\mathcal{O}_{F_w}) \operatorname{diag}(1, 1, p^{-1}, p^{-1}) \mathrm{GSp}_4(\mathcal{O}_{F_w})]$ (see [FP21, Rem. 5.6] for instance) which differs by an element of the centre from the spherical Hecke operator considered in §2.4.7. We justify this discrepancy as follows: when doing geometry and working with the moduli interpretation, we prefer to use this Hecke operator, while when doing local representation theory and considering Galois representations, we prefer to use the Hecke operators considered in §2.4. In this paper we will systematically work on spaces with fixed central character so that the (normalized) action of these two Hecke operators are the same. The same remark will apply to all the Hecke operators at p considered in this paper. We hope this will not cause any confusion.

3.9.20. *Hecke operators at p: Klingen type operator.* We again assume that $K = K^p K_p$ and $K_p = G_1(\mathbf{Z}_p)$. Let us fix a place w above p. We are going to define an action of a Hecke operator T_w on $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$, $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa)$ and their cuspidal versions. As before, the action on $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,\Sigma}, \omega^\kappa(-D))$ is not canonical and depends on the choice of x_w made in §3.6.3, but the action on $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,\Sigma}^{G_1}, \omega^\kappa(-D))$ is canonical.

Remark 3.9.21. The Hecke operator that we define in this section does not correspond to the double coset operator $[\mathrm{GSp}_4(\mathcal{O}_{F_w})\,\mathrm{diag}(1,p^{-1},p^{-1},p^{-2})\,\mathrm{GSp}_4(\mathcal{O}_{F_w})]$ but rather some variant of it that we call T_w. The formula for T_w in terms of double cosets is

$$\begin{aligned} T_w &= [\mathrm{GSp}_4(\mathcal{O}_{F_w})\,\mathrm{diag}(1,p^{-1},p^{-1},p^{-1})\mathrm{Par}(w)][\mathrm{Par}(w)\,\mathrm{diag}(1,1,1,p^{-1})\,\mathrm{GSp}_4(\mathcal{O}_{F_w})] \\ &= p[\mathrm{GSp}_4(\mathcal{O}_{F_w})\,\mathrm{diag}(1,p^{-1},p^{-1},p^{-2})\,\mathrm{GSp}_4(\mathcal{O}_{F_w})] \\ &\qquad + (1+p+p^2+p^3)p^{-2}[\mathrm{GSp}_4(\mathcal{O}_{F_w})\,\mathrm{diag}(p^{-1},p^{-1},p^{-1},p^{-1})\,\mathrm{GSp}_4(\mathcal{O}_{F_w})]. \end{aligned}$$

Set $K' = K^pK'_p$ where $K'_p = \prod_{v\neq w}\mathrm{GSp}_4(\mathcal{O}_{F_v})\times\mathrm{Kli}(w)$ and $K'' = K^pK''_p$ where $K''_p = \prod_{v\neq w}\mathrm{GSp}_4(\mathcal{O}_{F_v})\times\mathrm{Par}(w)$. In §3.6.3, we defined morphisms $p_1 : X_{K'} \to X_K$, $p_2 : X_{K'} \to X_{K''}$ giving a Hecke correspondence:

$$\begin{tikzcd} & X_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_{K''} & & X_K \end{tikzcd}$$

The key geometric properties are again (see Proposition 3.4.1):

(1) $X_{K'}$, $X_{K''}$ and X_K are relative complete intersections over $\mathrm{Spec}\,\mathbf{Z}_p$ of the same (pure) dimension,
(2) X_K is smooth over $\mathrm{Spec}\,\mathbf{Z}_p$,
(3) $X_{K'}$ and $X_{K''}$ are smooth over $\mathrm{Spec}\,\mathbf{Z}_p$ up to codimension 2 and normal.

We are again in the lci situation, so we have invertible dualizing sheaves $p_1^!\mathcal{O}_{X_K}$ and $p_2^!\mathcal{O}_{X_{K''}}$ and fundamental classes $p_1^*\mathcal{O}_{X_K} \to p_1^!\mathcal{O}_{X_K}$ and $p_2^*\mathcal{O}_{X_{K''}} \to p_2^!\mathcal{O}_{X_{K''}}$.

For all weights $\kappa = (k_v, l_v)$ with $l_v \geq 0$, we have natural maps $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ provided by the differential of the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ on $X_{K'}$, and $p_1^*\omega^\kappa \to p_2^*\omega^\kappa$ provided by the differential of the isogeny $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$. We therefore obtain two cohomological correspondences $\Theta_1 : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$ and $\Theta_2 : p_1^*\omega^\kappa \to p_2^!\omega^\kappa$.

Lemma 3.9.22. *When $l_w \geq 2$, the map Θ_1 is divisible by p^{2+l_w} and the map Θ_2 is divisible by p.*

Proof. This can be proved in exactly the same way as Lemma 3.9.18, by an explicit check over the completed local rings of generic points of the special fibre of $X_{K'}$. The details may be found in the proofs of [Pil20, Lem. 7.1.1, 7.1.2]. □

We can therefore consider the normalized fundamental classes $T'_w = p^{-2-l_w}\Theta_1 : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$ and $T''_w = p^{-1}\Theta_2 : p_1^*\omega^\kappa \to p_2^!\omega^\kappa$, and we obtain Hecke operators:

$$T'_w : \mathrm{R}\Gamma(X_{K''},\omega^\kappa) \to \mathrm{R}\Gamma(X_K,\omega^\kappa)$$

and

$$T''_w : \mathrm{R}\Gamma(X_K,\omega^\kappa) \to \mathrm{R}\Gamma(X_{K''},\omega^\kappa).$$

We set $T_w := T'_w \circ T''_w$. Similar definitions apply to cuspidal cohomology and work over $X_K^{G_1}$.

Remark 3.9.23. Just as the complexes that we are considering are independent of the choice of compactification by Lemma 3.8.2, so too are the actions of $T_{w,1}$ and T_w on them. See [Pil20, Prop. 7.2.1] for the case of T_w; the argument for $T_{w,1}$ is similar, but easier, and is left to the interested reader.

3.10. **Cohomology and automorphic representations.** Let $K = \prod_v K_v \subset \mathrm{GSp}_4(\mathbf{A}_F^\infty) = G_1(\mathbf{A}_{\mathbf{Q}}^\infty)$ be an open compact subgroup, let $S \supset S_p$ be a finite set of places such that $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ for $v \notin S$, and let

$$\widetilde{\mathbf{T}} = \bigotimes_{v \notin S} \mathcal{O}[\mathrm{GSp}_4(F_v) /\!/ \mathrm{GSp}_4(\mathcal{O}_{F_v})]$$

be the ring of spherical Hecke operators away from S. We say that a maximal ideal $\mathfrak{m} \subset \widetilde{\mathbf{T}}$ is *non-Eisenstein* if the residue field $\widetilde{\mathbf{T}}/\mathfrak{m}$ is a finite extension of $\mathbf{F}_p$, and for (any) inclusion $\widetilde{\mathbf{T}}/\mathfrak{m} \to \overline{\mathbf{F}}_p$ there exists an irreducible representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{F}}_p)$ with the property that, for each $v \notin S$, we have $\det(X - \overline{\rho}(\mathrm{Frob}_v))$

$$\equiv X^4 - T_{v,1}X^3 + (q_v T_{v,2} + (q_v^3 + q_v)T_{v,0})X^2 - q_v^3 T_{v,0} T_{v,1} X + q_v^6 T_{v,0}^2 \pmod{\mathfrak{m}}.$$

(cf. (2.4.8)).

Our main aim in this section is to prove the following result.

Theorem 3.10.1. *Let $\kappa = (k_v, l_v)_{v|\infty}$ with $k_v \geq l_v \geq 2$ and $k_v \equiv l_v \pmod 2$ be a weight and let $\mathfrak{m}$ be non-Eisenstein.*

(1) For $i = 0, 1$, there is an $\overline{E}[\mathrm{GSp}_4(\mathbf{A}_F^\infty)/\!/K]$-equivariant inclusion

$$\bigoplus_{\pi} (\pi^\infty)^K_{\mathfrak{m}} \otimes \overline{E} \subseteq H^i(X_K^{G_1}, \omega^\kappa(-D))^{|\cdot|^2}_{\mathfrak{m}} \otimes \overline{E} \tag{3.10.2}$$

where, on the right hand side, the superscript $|\cdot|^2$ indicates the space on which the diamond operators at places $v \notin S$ act via $|\cdot|^2$; and on the left hand side, π runs over the cuspidal automorphic representations of $\mathrm{GSp}_4(\mathbf{A}_F)$ with weight κ and central character $|\cdot|^2$ such that

- *π_v is holomorphic for those $v|\infty$ for which $l_v > 2$, and*
- *$\#\{v|\infty \mid \pi_v \text{ is not holomorphic}\} = i$.*

(2) There is an absolute constant R such that if for each $v|\infty$

- *$k_v - l_v > R$, and*
- *either $l_v = 2$ or $l_v > R$,*

then the inclusion (3.10.2) is an equality.

(3) If $i = 0$, then (3.10.2) is an equality. In fact, a version of this statement holds without having to localize at a non-Eisenstein maximal ideal; there is an $\overline{E}[\mathrm{GSp}_4(\mathbf{A}_F^\infty)/\!/K]$-equivariant isomorphism

$$H^0(X_K^{G_1}, \omega^\kappa(-D))^{|\cdot|^2} \otimes \overline{E} = \bigoplus_{\pi} (\pi^\infty)^K \tag{3.10.3}$$

where π runs over the cuspidal automorphic representations of $\mathrm{GSp}_4(\mathbf{A}_F)$ with weight κ and central character $|\cdot|^2$ which are holomorphic at all infinite places.

Remark 3.10.4. Theorem 3.10.1 is by no means optimal; the same results should hold for any cohomological degree i, and with a much weaker regularity assumption on κ in part (2). However, it seems difficult to deduce results in this generality from the literature, so we have restricted ourselves to this result, for which we only need to consider the cohomology of the boundary in degree 0. We explain the proof below, after proving a corollary and a preparatory lemma.

Theorem 3.10.1 has the following useful corollary.

Corollary 3.10.5. *Suppose that we are in the setting of Theorem 3.10.1 and the hypothesis on κ in (2) holds. Let l_0 denote the number of infinite places v with $l_v = 2$. Then*

$$\dim_{\overline{E}} H^1(X_K^{G_1}, \omega^\kappa(-D))_{\mathfrak{m}}^{|\cdot|^2} \otimes \overline{E} = l_0 \dim_{\overline{E}} H^0(X_K^{G_1}, \omega^\kappa(-D))_{\mathfrak{m}}^{|\cdot|^2} \otimes \overline{E}.$$

Proof. Since $\mathfrak{m}$ is non-Eisenstein, the automorphic representations π which contribute to (3.10.2) are all of general type in the sense of [Art04] by Lemma 2.9.1. There are l_0 ways to choose an infinite place v with $l_v = 2$, and we let π_v be generic for this place and holomorphic at the other infinite places. The result then follows from Theorem 2.9.3. □

Remark 3.10.6. The following lemma is essentially a special case of the much more general results proved in [HZ01], and can presumably be proved using the techniques of that paper, but since our Shimura varieties do not satisfy the precise assumptions needed to cite the results of [HZ01], we have chosen to give a direct proof.

Lemma 3.10.7. *Let $\kappa = (k_v, l_v)_{v|\infty}$ be a weight, with $k_v \geq l_v \geq 2$ and $k_v \equiv l_v \pmod 2$, and let $\mathfrak{m}$ be a non-Eisenstein maximal ideal. Let D denote the boundary of $X_K^{G_1}$. Then $H^0(D, \omega^\kappa)_{\mathfrak{m}} \otimes \overline{E} = 0$.*

Proof. In the case that $F = \mathbf{Q}$ this follows from [Fre83, IV, Satz 4.4], as in the proof of [Pil20, Cor. 15.2.3.1], so we can and do assume that $F \neq \mathbf{Q}$ in what follows. We let $\pi : X_K^{G_1} \to X_K^{*,G_1}$ be the map between toroidal and minimal compactifications. We let $\partial X^* \subset X_K^{*,G_1}$ be the (reduced) boundary of the minimal compactification, which we can write as $\partial_0 X^* \coprod \partial_1 X^*$, where $\partial_1 X^*$ is a union of Hilbert modular varieties for the group $\mathrm{Res}_{F/\mathbf{Q}}\mathrm{GL}_2$, and the complement $\partial_0 X^*$ is a finite union of points.

Suppose firstly that we are in the case that $k_v = l_v = k$ for some k independent of v. Then ω^κ is pulled back from the minimal compactification, and since $\pi_*(\omega^\kappa|_D) = \omega^\kappa|_{\partial X^*}$, we have $H^0(D, \omega^\kappa) = H^0(\partial X^*, \omega^\kappa)$. (To see that we have an identification $\pi_*(\omega^\kappa|_D) = \omega^\kappa|_{\partial X^*}$, it suffices by the projection formula to show that $\pi_*(\mathcal{O}_D) = \mathcal{O}_{\partial X^*}$. This follows from the facts that $\pi_*\mathcal{O}_{X_K^{G_1}} = \mathcal{O}_{X_K^{*,G_1}}$, $\pi_*\mathcal{I}_D = \mathcal{I}_{\partial X^*}$, and $R^1\pi_*\mathcal{I}_D = 0$.)

Suppose now that we are not in the case that k_v and l_v are equal and independent of v. Then it follows from the results of [Lan13], see [BR16, Prop. 1.5.8], that any element of $H^0(D, \omega^\kappa)$ vanishes on $\pi^{-1}(\partial_0 X^*)$; so the map $H^0(D, \omega^\kappa) \to H^0(\pi^{-1}(\partial_1 X^*), \omega^\kappa)$ is injective, and it suffices to show that $H^0(\pi^{-1}(\partial_1 X^*), \omega^\kappa)$ is Eisenstein. Again by [BR16, Prop. 1.5.8] it follows that $\pi_*(\omega^\kappa|_{\pi^{-1}(\partial_1 X^*)})$ is zero if the l_v are not all equal, and otherwise is equal to the sheaf $\omega^{(k_v)_{v|p}}$ (where we are using the usual labelling of weights for sheaves on Hilbert–Blumenthal modular schemes).

In either case, we have seen that the space that we are considering either vanishes, or injects into $H^0(\partial_1 X^*, \omega^{(k_v)})$. Now it is convenient to work adelically. Let us fix $W \in \mathfrak{C}$ with $\dim_F W = 1$. Then $\partial_1 X^*$ is as follows (where P_W and M_W are defined in §3.5):

$$\begin{aligned} &P_W^+(\mathbf{Q})\backslash \mathcal{H}_1^{[F:\mathbf{Q}]} \times G_1(\mathbf{A}^\infty)/K \\ &= P_W^+(\mathbf{Q})\backslash \mathcal{H}_1^{[F:\mathbf{Q}]} \times P_W(\mathbf{A}^\infty) \times^{P_W(\mathbf{A}^\infty)} G_1(\mathbf{A}^\infty)/K \\ &= M_W^+(\mathbf{Q})\backslash \mathcal{H}_1^{[F:\mathbf{Q}]} \times M_W(\mathbf{A}^\infty) \times^{P_W(\mathbf{A}^\infty)} G_1(\mathbf{A}^\infty)/K. \end{aligned}$$

We therefore find that

$$H^0(\partial_1 X^*, \pi_* \omega^{(k_v)}) = \left(\mathrm{Ind}_{P_W(\mathbf{A}^\infty)}^{G_1(\mathbf{A}^\infty)}(M)\right)^K$$

where $M = \lim_{K \subset M_W(\mathbf{A}^\infty)} H^0(M_W^+(\mathbf{Q})\backslash \mathcal{H}_1^{[F:\mathbf{Q}]} \times M_W(\mathbf{A}^\infty)/K, \omega^{(k_v)})$ from which it follows that the eigensystems arising from $H^0(\partial_1 X^*, \omega^{(k_v)})$ are Eisenstein. Indeed, since we are assuming that $F \neq \mathbf{Q}$, it follows from Koecher's principle that the cohomology groups $H^0(M_W^+(\mathbf{Q})\backslash \mathcal{H}_1^{[F:\mathbf{Q}]} \times M_W(\mathbf{A}^\infty)/K, \omega^{(k_v)})$ are spaces of Hilbert modular forms, and thus have associated two-dimensional Galois representations. More precisely, we have Satake transforms between spherical algebras (say at some unramified place v):

$$\mathcal{H}_{G_1} \to \mathcal{H}_{M_W} \to \mathbf{C}[X_*(T)]$$

for which the element $[1, 1, \varpi_v, \varpi_v]$ is mapped to

$$[1, \varpi_v, 1, \varpi_v] + [\varpi_v, 1, \varpi_v, 1] + [\varpi_v, \varpi_v, 1, 1] + [1, \varpi_v, 1, \varpi_v]$$
$$= ([1, \varpi_v, 1, \varpi_v] + [1, \varpi_v, 1, \varpi_v]) + [\varpi_v, 1, 1, \varpi_v^{-1}]([1, \varpi_v, 1, \varpi_v] + [1, \varpi_v, 1, \varpi_v]).$$

This expresses the relation between the Hecke operators on G_1 and M_W, so that if $\chi \times \pi$ is an automorphic representation contributing to M, it will contribute to the cohomology of G_1 via the compatible system of representations $\rho_\pi \oplus (\rho_\pi \otimes \chi)$. □

Before proving Theorem 3.10.1, we introduce some notation. Let

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

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

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

Let K^h denote the centralizer of h in $\mathrm{GSp}_4(\mathbf{R})$ (acting by conjugation). Then since $h(i) = J$, we see that we may identify $K^h = \mathbf{R}^\times U(2)$, so that $U(2)$ is a maximal compact subgroup of the identity component of $\mathrm{GSp}_4(\mathbf{R})$. Let $\mathfrak{g}_{\mathbf{C}} = \mathfrak{g}^{0,0} \oplus \mathfrak{g}^{-1,1} \oplus \mathfrak{g}^{1,-1}$ denote the Hodge structure on $\mathfrak{g}$, where $\mathfrak{g}^{0,0} = \mathfrak{k}_{h,\mathbf{C}}$ is the complex Lie algebra of K^h. Let $\mathfrak{p}^+ = \mathfrak{g}^{-1,1}$, $\mathfrak{p}^- = \mathfrak{g}^{1,-1}$, and $\mathfrak{P}_h = \mathfrak{k}_{h,\mathbf{C}} \oplus \mathfrak{p}^-$. We now define P^- to be the parabolic with Lie algebra $\mathfrak{P}_h$ with $P^- \cap \mathrm{GSp}_4(\mathbf{R}) = K^h$. We warn the reader that this parabolic is denoted by $\mathcal{P}_h$ in [Har90a], by Q^- in [BHR94], and by Q in [CG18]. (Note, however, that the fundamental object is really the Lie algebra $\mathfrak{P}_h$ because P^- only intervenes below via its Lie algebra.)

For each place $v|\infty$, we write P_v^- and K_v^h for the corresponding groups for $\mathrm{GSp}_4(F_v) \cong \mathrm{GSp}_4(\mathbf{R})$. We write V_{κ_v} for the representation of $K_v^h \cong \mathbf{R}^\times U(2)$ such that the automorphic vector bundle corresponding to $V_\kappa := \otimes_{v|\infty} V_{\kappa_v}$ via Definition 1.3.2 of [BHR94] is identified with ω^κ. We set $P_\infty^- := \prod_{v|\infty} P_v^-$ and $K_\infty^h := \prod_{v|\infty} K_v^h$.

Proof of Theorem 3.10.1. We begin by proving (3.10.2). By Lemma 3.10.7 we can and do replace $H^i(X_K^{G_1}, \omega^\kappa(-D))$ by the interior cohomology

$$\overline{H}^i(X_K^{G_1}, \omega^\kappa) := \mathrm{im}\left(H^i(X_K^{G_1}, \omega^\kappa(-D)) \to H^i(X_K^{G_1}, \omega^\kappa)\right).$$

(This is the only place that we use our assumption that $i \leq 1$, or the non-Eisenstein localization.) Let $\mathcal{A}_{\mathrm{cusp}}(G_1) \subset \mathcal{A}_{(2)}(G_1)$ be respectively the space of cuspidal automorphic forms on G_1 with central character $|\cdot|^2$, and the space of square integrable forms with this central character. By [Har90a, Thm. 2.7], we have inclusions

$$\bigoplus_{\pi\in\mathcal{A}_{\mathrm{cusp}}(G_1)} \Big((\pi^\infty)^K \otimes H^i(\mathrm{Lie}\, P_\infty^-, K_\infty^h; \pi_\infty \otimes V_\kappa)\Big)^{\oplus m_{\mathrm{cusp}}(\pi)} \subseteq \overline{H}^i(X_K^{G_1}, \omega_{\mathbf{C}}^\kappa)^{|\cdot|^2}$$

and

$$\overline{H}^i(X_K^{G_1}, \omega_{\mathbf{C}}^\kappa)^{|\cdot|^2} \subseteq \bigoplus_{\pi\in\mathcal{A}_{(2)}(G_1)} \Big((\pi^\infty)^K \otimes H^i(\mathrm{Lie}\, P_\infty^-, K_\infty^h; \pi_\infty \otimes V_\kappa)\Big)^{\oplus m_{(2)}(\pi)}$$

where $m_*(\pi)$ denotes the multiplicity of π in $\mathcal{A}_*(G_1)$. By Arthur's multiplicity formula for GSp_4 [Art04, GT19], we in fact have $m_{\mathrm{cusp}}(\pi) = m_{(2)}(\pi) = 1$ for all π.

In the proof of [BHR94, Thm. 4.2.3], it is shown that if $\pi \in \mathcal{A}_{(2)}(G_1)$ with $H^i(\mathrm{Lie}\, P_\infty^-, K_\infty^h; \pi_\infty \otimes V_\kappa) \neq 0$ and if the infinitesimal character of π_∞ is sufficiently far away from all the root hyperplanes that it does not lie on, then π_∞ is essentially tempered. In view of the relation between the infinitesimal character of π_∞ and κ arising from the Casselman–Osborne theorem (see [BHR94, Prop. 2.4.5]), the regularity condition on the infinitesimal character is exactly what we have assumed on κ in (2). Then by [Wal84, Thm. 4.3], we in fact have $\pi \in \mathcal{A}_{\mathrm{cusp}}(G_1)$, and so the inclusions above are equalities. In addition, by [Har90a, Thm. 3.5] (a theorem of Mirković) and [BHR94, Thm. 3.2.1], for each $v|\infty$ we have that $H^j(\mathrm{Lie}\, P_v^-, K_v^h; \pi_v \otimes V_{\kappa_v}) = 0$ unless either:

- $l_v > 2$, $j = 0$, and π_v is the holomorphic discrete series of weight (k_v, l_v), or;
- $l_v = 2$, $j = 0$, and π_v is the holomorphic limit of discrete series of weight (k_v, l_v), or;
- $l_v = 2$, $j = 1$, and π_v is the generic limit of discrete series of weight (k_v, l_v).

Moreover, in each of these cases that $H^j(\mathrm{Lie}\, P_v^-, K_v^h; \pi_v \otimes V_{\kappa_v})$ is nonzero, it is one-dimensional. The first two parts of the theorem then follow from the Künneth formula.

We now prove (3.10.3). In this case the map from $H^0(X_K^{G_1}, \omega^\kappa(-D))$ to the interior cohomology is an isomorphism by definition. Furthermore, by [Har90a, Prop. 2.7.2], the only π that contribute are automatically cuspidal (without needing to assume any regularity conditions). It follows from the theory of lowest weight representations, see for example [PS09, §2.3], that if $H^0(\mathrm{Lie}\, P_v^-, K_v^h; \pi_v \otimes V_{\kappa_v}) \neq 0$, then π_v is the holomorphic (limit of) discrete series of weight (k_v, l_v), as required. □

4. Hida complexes

In this section, we construct (higher) Hida theories for $\mathrm{GSp}_4(\mathbf{A}_F)$. The classical Hida theory is developed in [Hid04] and takes the form of a projective module over the total weight space (which is $2[F:\mathbf{Q}]$-dimensional). The construction of higher Hida theory was carried out when $F = \mathbf{Q}$ in [Pil20], and takes the shape of a perfect complex of amplitude $[0,1]$ over a one dimensional hyperplane of the weight space.

We assume that p splits completely in F and we construct all possible Hida theories, allowing the weight space at each place above p to be either 1- or 2-dimensional. Many of our arguments are simply the "product over the places $v|p$" of the arguments of [Hid04] and [Pil20]. To keep this paper at a reasonable length, we will often refer to [Pil20] for the details of arguments which go over directly to our case.

The bookkeeping needed to deal with having multiple places above p is considerable, and in the hope of orienting the reader, we begin this section with an overview of the arguments we will make. The main theorem of this section (and the only theorem that we will need later in the paper) is Theorem 4.6.1, which proves the existence of integral Hida complexes, and gives a control theorem for them in sufficiently high weight. Say that a classical weight $\kappa = (k_v, l_v)_{v|p}$ with $k_v \geq l_v \geq 2$ is "singular" at v if $l_v = 2$, and "regular" at v if $l_v > 2$. Fix any set I of places above p; these will be the places at which we interpolate automorphic forms of singular weight, while at the places in I^c, we interpolate forms of regular weight. (Thus traditional Hida theory considers the case $I = \emptyset$, while the higher Hida theory of [Pil20] is the case $F = \mathbf{Q}$ and $I = \{p\}$.)

There is a Hecke operator U^I (an analogue of the U_p operator for elliptic modular forms), which acts locally finitely on a complex of p-adic automorphic forms. The U^I-ordinary part M_I of this complex is a perfect complex over a weight space Λ_I, concentrated in degrees $[0, \#I]$. Furthermore, there is a constant C such that if $k_v - l_v \geq C$ and $l_v = 2$ for $v \in I$, and $l_v \geq C$ for $v \notin I$, then the H^0 of the specialization of M_I in weight κ agrees with the ordinary part of the degree 0 cohomology of $X_K^{G_1}$. (We expect that in fact this specialization should be quasi-isomorphic to the ordinary part of the classical cohomology, but we do not prove this. We do prove that there is also an injection of H^1s from the classical cohomology into that of M_I, which we will make use of in §6.)

The definition of M_I is motivated by the traditional case $I = \emptyset$ considered in [Hid04]. In that case one considers the cohomology at infinite Iwahori level over the ordinary locus, with coefficients in a certain interpolation sheaf which can be thought of as an interpolation of the highest weight vectors in the finite dimensional representations of the group GL_2/F. Since the ordinary locus is affine in the minimal compactification, one can prove that there is only cohomology in degree 0. Then one cuts out the ordinary part using a projector attached to U^I and proves that this defines a finite projective module over the Iwasawa algebra.

For general I, we instead consider the cohomology at infinite Klingen level of the locus which has p-rank at least 1 at places $w \in I$, and infinite Iwahori level over the ordinary (that is, p-rank 2) locus at places $w \in I^c$. This locus is no longer affine and it has cohomology in higher degrees. In fact by relating the cohomology of the toroidal and minimal compactifications, one can show that the cohomology is supported in degrees $[0, \#I]$.

One of the major difficulties in the proof of Theorem 4.6.1 is to show that the operator U^I acts locally finitely (in order to be able to associate an ordinary projector) and that the ordinary projection defines a perfect complex. By Nakayama's lemma for complexes, one reduces to showing that U^I has these properties for the cohomology modulo p, in some fixed weight. In particular, it suffices to consider the case of sufficiently large weight, i.e. the case that $k_v - l_v \geq C$ and $l_v = 2$ for $v \in I$, and $l_v \geq C$ for $v \notin I$, for some constant C. The first part of the argument is to relate this cohomology with the cohomology of the automorphic vector bundle of the corresponding weight over the locus $X_{K,1}^{G_1,I}$ of the special fibre of the Shimura variety which has p-rank at least 1 at the places $w \in I$, and is ordinary at places $w \in I^c$. This boils down to a computation at the level of the sheaf ω^κ itself, and to a computation in the Hecke algebra to show that the U^I-operator decreases the Klingen and Iwahori level.

In the case of Hida theory for 0-dimensional "Shimura varieties" (e.g. p-adic families of automorphic forms on definite unitary groups, as considered in [Che04, Ger19]) these arguments at the level of the sheaf and the Hecke algebra are all that is needed. In the geometric setting, more work is needed to establish the required finiteness of the ordinary part of the cohomology in characteristic p; recall that we are considering the cohomology on the locus $X_{K,1}^{G_1,I}$, so the cohomology groups are infinite dimensional before taking the ordinary parts. One has to show that (in sufficiently large weight, in characteristic p) ordinary cohomology classes on this locus extend to the whole Shimura variety (which is proper and has finite cohomology).

In order to do this, one shows that (again, in sufficiently regular weight, in characteristic p) the Hecke operator U^I acts by zero on the complement of $X_{K,1}^{G_1,I}$, so that after passing to ordinary parts, the cohomology agrees with that of the full Shimura variety, and is in particular finite-dimensional.

The vanishing of the Hecke operators on the part of the Shimura variety which is either of p-rank 0 (if $w \in I$) or is non-ordinary is accomplished by local calculations, using the definitions of the Hecke operators as cohomological correspondences. The case of $w \in I^c$ is relatively straightforward, as we are able to use the Hecke operator $T_{w,1}$ to prove this vanishing. (Note though that in this case we need to use the operator $U_{w,2}$, which is the operator at Klingen level corresponding to T_w, in the part of the argument explained above which takes place at the level of the sheaf.) The case $w \in I$ is much more delicate, as we need to use the operator T_w, which is significantly harder to control. (In this case, though, we use the same Hecke operators in the argument at the level of the sheaf as we do for the geometric part of the argument.)

The arguments below are in fact written in roughly the reverse order of the explanation above. We begin in §4.1 by recalling some standard results on Hasse invariants and the p-rank stratification, before proving the vanishing of the Hecke operators in small p-rank at spherical level in §4.2. In §4.3 and 4.4 we introduce the Igusa tower over the Shimura variety at Klingen and Iwahori level, and define the interpolation sheaves whose cohomology we use to define M_I. We then define the Hecke operator U^I in §4.5, and in §4.6 we prove Theorem 4.6.1, by relating the ordinary parts of the cohomology at spherical and Klingen level, and then carrying out the argument sketched above.

4.1. **Mod p-geometry: Hasse invariants and stratifications.** In this section, we introduce the p-rank stratification on our Siegel variety and the definition of several Hasse invariants attached to this stratification. The discussion follows [Pil20, §6.3, 6.4].

4.1.1. *Over X.* We assume that $K = K^p K_p$, $K_p = \prod_{v|p} K_v$ with

$$K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v)\}.$$

We fix a polyhedral cone decomposition Σ, and write $X = X_{K,\Sigma}$ if the context is clear. We let $\mathcal{G} = A[p^\infty]$ be the p-divisible group corresponding to the semi-abelian scheme A defined over X (well defined up to prime-to-p quasi-isogeny). This p-divisible group decomposes as $\mathcal{G} = \prod_{v|p} \mathcal{G}_v$. If $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$, the p-divisible group $\mathcal{G}_v$ defined over X carries a principal quasi-polarization. If $K_v = \mathrm{Par}(v)$ then the p-divisible group $\mathcal{G}_v$ carries a quasi-polarization of degree p^2: $\mathcal{G}_v \to \mathcal{G}_v^D$. Let

X_1 be the reduction of X modulo p. Then we let

$$\mathrm{Ha}(\mathcal{G}_v) \in H^0(X_1, \det \omega_{\mathcal{G},v}^{p-1})$$

be the Hasse invariant corresponding to $\mathcal{G}_v$; it is compatible with étale isogenies (by construction) and also with duality.

For any place $v|p$, we let $X_1^{\geq_v 2} = X_1^{=_v 2}$ be the open subscheme defined by $\mathrm{Ha}(\mathcal{G}_v) \neq 0$. This is the ordinary locus at v. We let $X_1^{\leq_v 1}$ be its complement defined by $\mathrm{Ha}(\mathcal{G}_v) = 0$. This is the non-ordinary locus at v. It carries the reduced schematic structure by the proof of [Pil20, Lem. 6.4.1]. (Whenever we use notation of the form $X_1^{\geq_v 2}$, $X_1^{\leq_v 1}$ etc., the superscript is referring to the multiplicative rank of the group scheme $\mathcal{G}_v$.)

As a very special case of the general constructions of [Box15, GK19], there is a secondary Hasse invariant

$$\mathrm{Ha}'(\mathcal{G}_v) \in H^0(X_1^{\leq_v 1}, \det \omega_{\mathcal{G},v}^{p^2-1})$$

(see also [Pil20, §6.3.2] when $K = \mathrm{Par}(v)$). Its non-vanishing locus is $X_1^{=_v 1}$, the rank 1 locus at v. We define its schematic complement $X_1^{=_v 0}$, the supersingular locus at v, by the equation $\mathrm{Ha}'(\mathcal{G}_v) = 0$. It carries a non-reduced schematic structure, see [Pil20, Rem. 6.4.1].

We can intersect the locally closed subschemes we have defined. Consider disjoint subsets $I_1, \dots, I_r \subset \{v|p\}$, symbols $*(i) \in \{\leq, \geq, =\}$ for $1 \leq i \leq r$ and numbers $a_i \in \{0, 1, 2\}$ for $1 \leq i \leq r$. Then we define $X_1^{*(i)_{I_i} a_i,\ i=1,\cdots,r}$ as the intersection of the spaces $X_1^{*(i)_v a_i}$ for all $1 \leq i \leq r$ and $v \in I_i$. It will be convenient to denote by $X_1^{\geq 2} = X_1^{\geq_{\{v|p\}} 2}$ the ordinary locus and by $X_1^{\geq 1} = X_1^{\geq_{\{v|p\}} 1}$ the rank 1 locus.

Note that for any disjoint sets I, J, K, the scheme $X_1^{\leq_I 1, \geq_J 1, \geq_K 2}$ is Cohen–Macaulay, and indeed is a local complete intersection over $\operatorname{Spec} \mathbf{F}_p$. To see this, note that $X_1^{\leq_I 1, \geq_J 1, \geq_K 2}$ is open in $X_1^{\leq_I 1}$, and $X_1^{\leq_I 1}$ is a complete intersection in X_1, because it is given by the vanishing of the Hasse invariants $\mathrm{Ha}(\mathcal{G}_v)$ for $v \in I$. Since X_1 itself is local complete intersection by Proposition 3.5.4, the result follows. We will in particular repeatedly use this fact in order to apply Lemma 3.8.10.

We will also frequently use some well-known results on the density of the ordinary locus, and on the density of the p-rank 1 locus in the p-rank less than or equal to 1 locus. We will need these results in slightly greater generality than has been considered above. To this end, consider disjoint subsets $I_1, \cdots, I_r$ as above, and let $v|p$ be a place not contained in $I_1 \cup \cdots \cup I_r$.

We assume that $K = K^p K_p$, $K_p = \prod_{w|p, w \neq v} K_w \times K_v$, and that

$$K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v), \mathrm{Kli}(v), \mathrm{Si}(v), \mathrm{Iw}(v)\},$$

while $K_w \in \{\mathrm{GSp}_4(\mathcal{O}_{F_w}), \mathrm{Par}(w)\}$ for $w \neq v$. We can define topological spaces

$$|X_{K,1}^{*(i)_{I_i} a_i}|,\ |X_{K,1}^{*(i)_{I_i} a_i, =_v 2}|,\ |X_{K,1}^{*(i)_{I_i} a_i, \leq_v 1}|,\ \text{and}\ |X_{K,1}^{*(i)_{I_i} a_i, =_v 1}|$$

using the p-rank stratification as before. The point is that the p-rank is invariant under isogeny so we can consider the p-rank of any of the Barsotti–Tate groups of the chain. Note that one could give these spaces a schematic structure by using the Hasse invariants, but this structure will in general depend on which Barsotti–Tate group of the chain we use to define the Hasse invariants.

Our claims about density are then the following: $|X_{K,1}^{*(i)_{I_i}a_i,=_v 2}|$ is dense in $|X_{K,1}^{*(i)_{I_i}a_i}|$, while if we further assume that $K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v), \mathrm{Kli}(v)\}$, $|X_{K,1}^{*(i)_{I_i}a_i,=_v 1}|$ is dense in $|X_{K,1}^{*(i)_{I_i}a_i,\leq_v 1}|$. To see this, it suffices to prove the first statement in the case $K_v = \mathrm{Iw}(v)$, and the second statement in the case $K_v = \mathrm{Kli}(v)$. It then suffices to prove the corresponding statements for the corresponding local models, which follows easily from an explicit calculation. Indeed, the first statement is already proved in [dJ93], while the second follows from an analysis of the Kottwitz–Rapoport stratification at Iwahori level, and its image at Klingen level; see [Yu11, Thm. 4.2] for a precise statement.

4.1.2. *Over X^{G_1}.* The p-rank stratification is independent of the polarization and therefore all of the spaces we have defined in this section carry an induced action of $(\mathcal{O}_F)_{(p)}^{\times,+}$. It follows that the stratification descends to a stratification on $X_1^{G_1}$. We moreover observe that the sheaf $\det \omega_{\mathcal{G},v}^{p-1}$ can be canonically descended to a sheaf $\det \omega_{\mathcal{G},v}^{p-1}$ on $X_1^{G_1}$ (see Remark 3.7.3). It follows that the Hasse invariants $\mathrm{Ha}(\mathcal{G}_v)$ and $\mathrm{Ha}'(\mathcal{G}_v)$ (whose definition is independent of the polarization) also descend to sections of this sheaf over $X_1^{G_1}$ and $X_1^{G_1,\leq_v 1}$ respectively.

Therefore, if we consider disjoint subsets $I_1, \cdots, I_r \subset \{v|p\}$, symbols $*(i) \in \{\leq, \geq, =\}$ for $1 \leq i \leq r$ and numbers $a_i \in \{0,1,2\}$ for $1 \leq i \leq r$, there is a unique locally closed subscheme $(X_{K,1}^{G_1})^{*(i)_{I_i}a_i,\ i=1,\cdots,r}$ of $X_{K,1}^{G_1}$ whose inverse image in $X_{K,1}$ is $(X_{K,1})^{*(i)_{I_i}a_i,\ i=1,\cdots,r}$.

Remark 4.1.3. In §4.3.4 below we will define some other locally closed subschemes of the special fibres of the spaces $X_{K,1}$ at Klingen and Iwahori level, which will be important in the rest of the paper. We caution the reader that these will *not* be defined in terms of the p-ranks of the $\mathcal{G}_v[p]$, but will rather depend on subschemes of $\mathcal{G}_v[p]$ given by the Klingen and Iwahori level structures.

4.2. **Vanishing theorem for ordinary cohomology.** We assume that $K_p = G_1(\mathbf{Z}_p)$. Let $\kappa = (k_v, l_v)_{v|p}$ be a weight (recall from §3.7.2 that we are assuming that it satisfies the parity condition $k_v - l_v = 0 \mod 2$). Let $S_p := \{v|p\} = I \coprod I^c$ be a partition. We write $X_1^I := X_1^{\geq_I 1, \geq_{I^c} 2} \hookrightarrow X_1$, an open subscheme, and similarly $X_1^{G_1,I} \hookrightarrow X_1^{G_1}$. The main theorem of this subsection is:

Theorem 4.2.1. *Let $T^I = \prod_{w\in I} T_w \prod_{w \in I^c} T_{w,1}$. There is a universal constant C depending only on p and F but not on the tame level K^p such that if $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w \in I$, and $l_w \geq C$ for all $w \in I^c$, then $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ carries a locally finite action of T^I. Furthermore, under this assumption on κ,*

(1) $e(T^I)\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ is a perfect complex of amplitude $[0, \#I]$.
(2) The map $e(T^I)H^0(X_1^{G_1}, \omega^\kappa(-D)) \to e(T^I)H^0(X_1^{G_1,I}, \omega^\kappa(-D))$ is an isomorphism.
(3) The map $e(T^I)H^1(X_1^{G_1}, \omega^\kappa(-D)) \to e(T^I)H^1(X_1^{G_1,I}, \omega^\kappa(-D))$ is injective.
(4) If furthermore $l_w \geq 3$ for all $w \in I$, then

$$e(T^I)\mathrm{R}\Gamma(X_1^{G_1}, \omega^\kappa(-D)) \to e(T^I)\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$$

is a quasi-isomorphism.

Here $e(T^I)$ is the ordinary projector associated to the operator T^I (see §2.11). We remark that $(2), (3), (4)$ of the theorem hold true for non-cuspidal cohomology as well.

Remark 4.2.2. Various improvements on Theorem 4.2.1 should be possible. For example, the reader will see from the proof below that it is possible to prove that the Hecke operators at each place act locally finitely (rather than just proving it for their product), provided they satisfy explicit mild bounds on the weights (rather than depending on the indeterminate constant C); see Remark 4.2.34 for one approach to this. It may also be possible to give explicit values of C. For the purposes of this paper the statement of Theorem 4.2.1 suffices, and is well-adapted to a (somewhat involved) inductive proof working one place at a time.

We now briefly explain the main idea of the proof. We will often work at the level of X_1 rather than $X_1^{G_1}$. It is easier to work on X_1 because of the moduli interpretation. One can always deduce results for the cohomology on $X_1^{G_1}$ from results on the cohomology for X_1 by Proposition 3.8.3. We nevertheless warn the reader that X_1 has infinitely many connected components and therefore one cannot expect any finiteness results for the cohomology over X_1; accordingly, we work over $X_1^{G_1}$ when we want to show that a Hecke operator acts locally finitely.

The basic principle underlying these arguments is that the ordinary projectors $e(T_{w,1})$, $e(T_w)$ can be used to kill many cohomology classes. This idea is already used in [Pil20, §7, §8] (this is what we call Klingen vanishing below, because the Hecke operator T_w is associated with the Klingen parabolic) and of course also in [Hid04] (this is what we call Siegel vanishing, because the Hecke operator $T_{w,1}$ is associated with the Siegel parabolic).

We will typically not comment on the commutativity of the actions of Hecke operators at one place with multiplication by Hasse invariants at other places, which is easily checked.

4.2.3. *Vanishing theorems: Siegel vanishing.* Let $K = K^p K_p$ be a reasonable level at p. We assume that $K_p = G_1(\mathbf{Z}_p)$. We let $X = X_{K,\Sigma}$ and $X^{G_1} = X_{K,\Sigma}^{G_1}$. Let $\kappa = (k_v, l_v)_{v|p}$ be a weight. We begin with the following theorem.

Theorem 4.2.4. *There is a universal constant C depending only on p and F but not on the tame level K^p such that if $J \subseteq S_p$, and for each $w \in J$, we have $l_w \geq C$, then $\mathrm{R}\Gamma(X_1^{G_1,\geq_J 2}, \omega^\kappa)$ has a locally finite action of $T^J := \prod_{w\in J} T_{w,1}$, and*

$$e(T^J)\mathrm{R}\Gamma(X_1^{G_1}, \omega^\kappa) \to e(T^J)\mathrm{R}\Gamma(X_1^{G_1,\geq_J 2}, \omega^\kappa)$$

is a quasi-isomorphism. In particular $e(T^J)\mathrm{R}\Gamma(X_1^{G_1,\geq_J 2}, \omega^\kappa)$ is a perfect complex. The analogous statements also hold for cuspidal cohomology.

Remark 4.2.5. In fact Theorem 4.2.4 is not quite strong enough for our purposes; we will later replace it with Theorem 4.2.13, which is proved in exactly the same way. Our justification for presenting the material in this way is that the proof of Theorem 4.2.4 is a good warmup for the arguments that we will later make to prove "Klingen vanishing", and it seems simplest to make these arguments before considering the Klingen level Hecke operators and the much more complicated statements and arguments that we make in that context.

We only give the proof in the non-cuspidal case. The arguments go through unchanged in the cuspidal setting. The proof of Theorem 4.2.4 is by induction on $\#J$

(the case $J = \emptyset$ being vacuous), and depends on several lemmas. In our inductive argument we will feel free to increase the constant C in a manner depending only on J without comment. Write $J = J' \cup \{w\}$, and assume that Theorem 4.2.4 holds for J'.

Recall from §3.9.17 that the correspondence underlying the operator $T_{w,1}$ is $X_{K'}$ with $K' = K^p K'_p$ and $K'_p = \prod_{v \neq w} K_v \times \mathrm{Si}(w)$. We let $(X_{K'})_1$ denote the special fibre of this correspondence. Let $\kappa = (k_v, l_v)_{v|p}$ be a weight such that $l_w \geq 2$, so that we have a cohomological correspondence $T_{w,1} : p_2^* \omega^\kappa \to p_1^! \omega^\kappa$. By reduction modulo p, it follows from Lemma 3.8.10 (and the flatness of $X_{K'}$ and X_K over $\mathbf{Z}_p$) that we get a cohomological correspondence still denoted $T_{w,1} : p_2^*(\omega^\kappa|_{X_1}) \to p_1^!(\omega^\kappa|_{X_1})$. This cohomological correspondence is a map of locally free sheaves over $(X_{K'})_1$. As in §3.8.13, this correspondence pulls back to the open subscheme $X_1^{G_1, \geq_{J'}2}$.

Adopting the notation of §4.1 we consider the dense open subscheme $X_{K',1}^{\geq_J 2} = (X_{K',1}^{\geq_{J'}2})^{=_w 2}$ of $X_{K',1}^{\geq_{J'}2}$, which is by definition the ordinary locus at w (that is, the locus for which $\mathcal{G}_w$ is ordinary). This scheme is the union of several types of connected components. Let $p_1^* \mathcal{G}_w \to p_2^* \mathcal{G}_w$ be the universal map on the p-divisible group. We let $(X_{K',1}^{\geq_{J'}2})^{=_w 2, et}$ be the étale components (that is, those for which the kernel of this isogeny doesn't contain a multiplicative group), and we let $(X_{K',1}^{\geq_{J'}2})^{=_w 2, net}$ be the other components. We can therefore decompose the cohomological correspondence $T_{w,1}$ over $(X_{K',1}^{\geq_{J'}2})^{=_w 2}$ into $T_{w,1} = T_{w,1}^{et} + T_{w,1}^{net}$ where $T_{w,1}^{et}$ is the projection of $T_{w,1}$ on the étale components and $T_{w,1}^{net}$ is the projection on the other components.

Lemma 4.2.6. *The map $T_{w,1}^{net}$ is zero as soon as $l_w \geq 3$. For all $l_w \geq 3$ we have a commutative diagram of maps of sheaves over $(X_{K',1}^{\geq_{J'}2})^{=_w 2}$:*

$$\begin{tikzcd}
p_2^* \omega^\kappa \arrow[r, "T_{w,1}^{et}"] \arrow[d, "p_2^* \mathrm{Ha}(\mathcal{G}_w)"] & p_1^! \omega^\kappa \arrow[d, "p_1^* \mathrm{Ha}(\mathcal{G}_w)"] \\
p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T_{w,1}^{et}"] & p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})
\end{tikzcd}$$

Proof. The first point follows from an inspection of the proof of Lemma 3.9.18 (since $l_w + 1, k_w + l_w > 3$). The second point follows from the fact that the Hasse invariant commutes with étale isogenies. □

Lemma 4.2.7. *The following diagram of locally free sheaves on $X_{K',1}^{\geq_{J'}2}$ is commutative for $l_w \geq 3$:*

$$\begin{tikzcd}
p_2^* \omega^\kappa \arrow[r, "T_{w,1}"] \arrow[d, "p_2^* \mathrm{Ha}(\mathcal{G}_w)"] & p_1^! \omega^\kappa \arrow[d, "p_1^* \mathrm{Ha}(\mathcal{G}_w)"] \\
p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T_{w,1}"] & p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})
\end{tikzcd}$$

Proof. Since $X_{K',1}^{\geq_{J'}2}$ is Cohen–Macaulay and all of the sheaves are locally free, it suffices to prove the commutativity over a dense open subscheme. We may therefore prove it over $(X_{K',1}^{\geq_{J'}2})^{=_w 2}$, so we are done by Lemma 4.2.6. □

Lemma 4.2.8. *$\mathrm{Ha}(\mathcal{G}_w)$ is not a zero divisor on $X_1^{\geq_{J'}2}$, and $p_2^*\mathrm{Ha}(\mathcal{G}_w)$ is not a zero divisor on $X_{K',1}^{\geq_{J'}2}$.*

Proof. Since $X_1^{\geq_{J'}2}$ and $X_{K',1}^{\geq_{J'}2}$ are Cohen–Macaulay, this follows from the fact that the non-ordinary loci have codimension 1. □

In what follows, we warn the reader that while the schemes $p_i^{-1}(X_1^{\geq_{J'}2,\leq_w 1}) = X_{K',1}^{\geq_{J'}2} \times_{p_i, X_1^{\geq_{J'}2}} X_1^{\geq_{J'}2,\leq_w 1}$ for $i = 1, 2$ have the same underlying topological spaces, they have different (non reduced) scheme structures. In particular sheaves like $p_i^*\omega^\kappa|_{X^{\geq_{J'}2,\leq_w 1}}$ for $i = 1, 2$ have different (scheme theoretic) support.

Lemma 4.2.9. *For all $l_w \geq p + 2$ the cohomological correspondence $T_{w,1}$ restricts to give a cohomological correspondence*

$$T_{w,1} : p_2^*(\omega^\kappa|_{X_1^{\geq_{J'}2,\leq_w 1}}) \to p_1^!(\omega^\kappa|_{X_1^{\geq_{J'}2,\leq_w 1}}).$$

Proof. Since the cokernel of

$$p_2^*\omega^\kappa \xrightarrow{p_2^*\mathrm{Ha}(\mathcal{G}_w)} p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})$$

is $p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}|_{X_1^{\geq_{J'}2,\leq_w 1}})$, and (by Lemmas 3.8.10 and 4.2.8) the cokernel of

$$p_1^!\omega^\kappa \xrightarrow{p_1^*\mathrm{Ha}(\mathcal{G}_w)} p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})$$

is $p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}|_{X_1^{\geq_{J'}2,\leq_w 1}})$, it follows from Lemma 4.2.7 that provided that $l_w \geq 3$, $T_{w,1}$ restricts to give a cohomological correspondence

$$p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}|_{X_1^{\geq_{J'}2,\leq_w 1}}) \to p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}|_{X_1^{\geq_{J'}2,\leq_w 1}}),$$

as required. □

Lemma 4.2.10. *There is a universal constant C which depends only on F and p (but not on the tame level K^p) such that the map of Lemma 4.2.9*

$$T_{w,1} : p_2^*(\omega^\kappa|_{X_1^{\geq_{J'}2,\leq_w 1}}) \to p_1^!(\omega^\kappa|_{X_1^{\geq_{J'}2,\leq_w 1}})$$

is zero for all $l_w \geq C$.

Proof. We may and do assume that $J' = \emptyset$, as the general case follows immediately from this by restriction to an open. We moreover note that it suffices to find such a constant C for a single tame level K^p. Indeed, if $K_1^p \subseteq K_2^p$ are two choices of tame level, the natural forgetful map $X_{K'_pK_1^p} \to X_{K'_pK_2^p}$ commutes with p_1 and p_2, and is faithfully flat, from which it follows that C works for K_1^p if and only if it works for K_2^p.

Let $\mathcal{J}$ be the ideal defining $X_1^{\leq_w 1}$ in X and let $\mathcal{I} = p_1^*\mathcal{J}$. We need to prove that for l_w sufficiently large, the cohomological correspondence over $X_{K'}$, $T_{w,1} : p_2^*\omega^\kappa \to p_1^!\omega^\kappa$ factors through $T_{w,1} : p_2^*\omega^\kappa \to \mathcal{I}p_1^!\omega^\kappa$.

By definition, we have $T_{w,1} = p^{-3}\Theta(\kappa)$, where $\Theta(\kappa)$ is the composite of a map $\Theta_1(\kappa) : p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ and a fundamental class $\Theta_2(\kappa) : p_1^*\omega_\kappa \to p_1^!\omega^\kappa$, so in turn we need to show that $\Theta(\kappa)$ factors through $p^3\mathcal{I}p_1^!\omega^\kappa$ (of course, we have already shown that it factors through $p^3p_1^!\omega^\kappa$).

Let x be a generic point of $V(\mathcal{I}) \subset X_{K'}$. It corresponds to a Barsotti–Tate group in characteristic p whose p-rank at w is exactly one. The map $p_2^* \det \omega_{\mathcal{G}_w} \to$

$p_1^* \det \omega_{\mathcal{G}_w}$ is zero over $k(x)$ because the isogeny $p_1^*\mathcal{G}_w \to p_2^*\mathcal{G}_w$ is not étale at x. Let $I(x)$ be the ideal defining the Zariski closure $\overline{x}$ in $X_{K'}$. We deduce that the map $\Theta_1(\kappa) : p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ factors through $I(x)^{l_w} p_1^*\omega^\kappa$.

It follows that the map $\Theta(\kappa)$ factors through $p^3 p_1^*\omega^\kappa \bigcap \cap_x I(x)^{l_w} p_1^*\omega^\kappa$. By the Artin–Rees lemma, it factors through $p^3\mathcal{I}p_1^*\omega^\kappa$ for l_w larger than a constant C, as required. (We note that strictly speaking $V(I)$ has infinitely many connected components, however there are only finitely many orbits for the action of $(\mathcal{O}_F)_{(p)}^{\times,+}$, so there is some constant C which works for all of them.) □

Proof of Theorem 4.2.4. Take C as in Lemma 4.2.10. Recall that we write $J = J' \cup \{w\}$, and we are assuming that the theorem holds for J'. We begin by showing that the action of T^J on $\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa)$ is locally finite. By the inductive hypothesis, the action of $T^{J'}$ on this complex is locally finite, and $e(T^{J'})\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa)$ is perfect, so that in particular the action of T^J on $e(T^{J'})\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa)$ is locally finite. It is therefore enough to show that T^J acts locally finitely on $(1 - e(T^{J'}))\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa)$. Since $T^{J'}$ acts locally nilpotently on the complex $(1 - e(T^{J'}))\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa)$ by definition, so does T^J, so in particular it acts locally finitely, as required.

Now we consider the exact triangle

$$\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa) \to \mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \to \mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \xrightarrow{+1}$$

The operator T^J acts everywhere and is zero on $\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})$ by Lemma 4.2.10. We therefore deduce that

$$e(T^J)\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa) = e(T^J)\mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}).$$

Since $\mathrm{R}\Gamma(X_1^{G_1,\geq_J 2}, \omega^\kappa) = \varinjlim_n \mathrm{R}\Gamma(X_1^{G_1,\geq_{J'}2}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p-1)})$, the theorem follows. □

4.2.11. *Vanishing theorem: Siegel and Klingen vanishing.* We now turn to the more general situation, which involves the study of the Hecke operator T_w. Our analysis is similar to that of §4.2.3, but it is rather more involved because T_w is defined as the composite of two correspondences and because we need to study the supersingular locus at w rather than the non-ordinary locus at w. This subsection is devoted to the proof of the following theorem, which implies most of Theorem 4.2.1.

Let I_a, I_b, J be pairwise disjoint subsets of S_p. Then we will write

$$X_1^{I_{a,b},J} := X_1^{\leq_{I_a}1,\geq_{I_b}1,\geq_J 2} \text{ and } X_1^{G_1,I_{a,b},J} := X_1^{G_1,\leq_{I_a}1,\geq_{I_b}1,\geq_J 2}.$$

Theorem 4.2.12. *Let I_a, I_b, J be pairwise disjoint subsets of S_p. Set $T^{I_b,J} = \prod_{w\in I_b} T_w \prod_{w\in J} T_{w,1}$. Then there is a universal constant C depending only on p and F but not on the tame level K^p such that if $k_w - l_w \geq C$ and $l_w \geq 2$ for all $w \in I_b$, and $l_w \geq C$ for all $w \in J$, then $\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$ carries a locally finite action of $T^{I_b,J}$. Furthermore:*

(1) $e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$ is a perfect complex.
(2) The map

$$e(T^{I_b,J})H^0(X_1^{G_1,\leq_{I_a}1}, \omega^\kappa) \to e(T^{I_b,J})H^0(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$$

is an isomorphism.

(3) The map

$$e(T^{I_b,J})H^1(X_1^{G_1,\leq_{I_a}1},\omega^\kappa) \to e(T^{I_b,J})H^1(X_1^{G_1,I_{a,b},J},\omega^\kappa)$$

is injective.

(4) If furthermore $l_w \geq 3$ for all $w \in I_b$, then

$$e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1},\omega^\kappa) \to e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa)$$

is a quasi-isomorphism.

Moreover, the same results hold for cuspidal cohomology.

We only give the proof in the non-cuspidal setting. The same arguments work in the cuspidal case. The proof of this result again depends on several lemmas. We will firstly prove the result in the case $I_b = \emptyset$, by induction on $\#J$. We will then prove the general case by induction on $\#I_b$.

Recall from §3.9.17, that if we set $K' = K^pK'_p$ with $K'_p = \prod_{v\neq w} K_v \times \mathrm{Si}(w)$, there is a cohomological correspondence of Siegel type:

$$T_{w,1} : p_2^*(\omega^\kappa|_{X_K}) \to p_1^!(\omega^\kappa|_{X_K}).$$

By reduction modulo p and Lemma 3.8.10, we again get a cohomological correspondence: $T_{w,1} : p_2^*(\omega^\kappa|_{X_{K,1}}) \to p_1^!(\omega^\kappa|_{X_{K,1}})$. We let $X_{K',1}^{I_{a,b},J}$ be the pre-image of $X_1^{I_{a,b},J}$ in $X_{K',1}$ (via any of the projections, it doesn't matter). These correspondences may be restricted to $X_{K',1}^{I_{a,b},J}$ whenever $w \notin I_a$ by another application of Lemma 3.8.10, because this correspondence obviously commutes with the Hasse invariants at places in I_a.

Theorem 4.2.13. *There is a universal constant C depending only on p and F but not on the tame level K^p such that if I_a, J are disjoint, and if $l_w \geq C$ for all $w \in J$, then $\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1,\geq_J 2},\omega^\kappa)$ has a locally finite action of $T^{\emptyset,J} := \prod_{w\in J} T_{w,1}$, and*

$$e(T^{\emptyset,J})\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1},\omega^\kappa) \to e(T^{\emptyset,J})\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1,\geq_J 2},\omega^\kappa)$$

is a quasi-isomorphism. In particular $e(T^{\emptyset,J})\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1,\geq_J 2},\omega^\kappa)$ is a perfect complex.

Proof. The case $I_a = \emptyset$ is Theorem 4.2.4, and the theorem at hand may be proved by an identical inductive argument on $\#J$, once we have proved the base case $J = \emptyset$. But in this case $X_1^{G_1,\leq_{I_a}1}$ is proper, so $\mathrm{R}\Gamma(X_1^{G_1,\leq_{I_a}1},\omega^\kappa)$ is a perfect complex, and we are done. □

We now reintroduce Klingen type correspondences. Let $w|p$. By §3.9.20 if we set $K' = K^pK'_p$ with $K'_p = \prod_{v\neq w} K_v \times \mathrm{Kli}(w)$, and $K'' = K^pK''_p$ with $K''_p = \prod_{v\neq w} K_v \times \mathrm{Par}(w)$, there are cohomological correspondences of Klingen type: $T'_w : p_2^*(\omega^\kappa|_{X_{K''}}) \to p_1^!(\omega^\kappa|_{X_K})$ and $T''_w : p_1^*(\omega^\kappa|_{X_K}) \to p_2^!(\omega^\kappa|_{X_{K'}})$ for all weights $\kappa = (k_v, l_v)$ with $k_w \geq l_w \geq 2$. By reduction modulo p and Lemma 3.8.10, we again get cohomological correspondences: $T'_w : p_2^*(\omega^\kappa|_{X_{K'',1}}) \to p_1^!(\omega^\kappa|_{X_{K,1}})$ and $T''_w : p_1^*(\omega^\kappa|_{X_{K,1}}) \to p_2^!(\omega^\kappa|_{X_{K'',1}})$. We let $X_{K',1}^{I_{a,b},J}$ be the pre-image of $X_1^{I_{a,b},J}$ in $X_{K',1}$. These correspondences may be restricted to $X_{K',1}^{I_{a,b},J}$ whenever $w \notin I_a$ by another application of Lemma 3.8.10, because they obviously commute with the Hasse invariants at places in I_a.

In the rest of this section we prove Theorem 4.2.12 by induction on $\#I_b$. To this end, choose $w \in I_b$, write $I = I' \coprod \{w\}$, and write $I'_a = I' \cap I_a = I_a$, and $I'_b = I' \cap I_b$. We assume that Theorem 4.2.12 holds (for some value of C, which we fix) for all smaller values of $\#I_b$ (as we may, having proved the case $I_b = \emptyset$ in Theorem 4.2.4).

We now consider the scheme $(X_{K',1}^{I'_{a,b},J})^{=_w 2}$ (which is again by definition the subscheme where $\mathcal{G}_w$ is ordinary), which decomposes into several components. Let $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ be the universal isogeny of degree p^3. We denote by $(X_{K',1}^{I'_{a,b},J})^{=_w 2,et}$ the "étale" components, namely those where the kernel of the universal isogeny has multiplicative rank 1 (so it is as étale as possible), and by $(X_{K',1}^{I'_{a,b},J})^{=_w 2,net}$ the other components where the kernel has multiplicative rank 2.

This provides a decomposition of the correspondence $T'_w = T'_{w,et} + T'_{w,net}$ where $T'_{w,et}$ stands for the projection on the "étale" components and $T'_{w,net}$ for the projection on the "non-étale" components.

We also have an isogeny $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ of degree p and over $(X_{K',1}^{I'_{a,b},J})^{=_w 2,et}$ this isogeny has multiplicative kernel, while it is étale over $(X_{K',1}^{I'_{a,b},J})^{=_w 2,net}$ (observe that the étale and non-étale components are interchanged when we pass from the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ to the isogeny $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$). This provides a second decomposition $T''_w = T''_{w,et} + T''_{w,net}$ where $T''_{w,et}$ stands for the projection on $(X_{K',1}^{I'_{a,b},J})^{=_w 2,net}$ and $T''_{w,net}$ for the projection on $(X_{K',1}^{I'_{a,b},J})^{=_w 2,et}$.

Lemma 4.2.14. *If $l_w \geq 2$, and $k_w \geq 3$ then over $(X_{K',1}^{I'_{a,b},J})^{=_w 2}$ we have $T'_{w,net} = T''_{w,net} = 0$. Moreover, the following diagrams are commutative:*

$$\begin{tikzcd}
p_2^*\omega^\kappa \arrow[r, "T'_{w,et}"] \arrow[d, "p_2^*\mathrm{Ha}(\mathcal{G}_w)"] & p_1^!\omega^\kappa \arrow[d, "p_1^*\mathrm{Ha}(\mathcal{G}_w)"] \\
p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T'_{w,et}"] & p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})
\end{tikzcd}$$

$$\begin{tikzcd}
p_1^*\omega^\kappa \arrow[r, "T''_{w,et}"] \arrow[d, "p_1^*\mathrm{Ha}(\mathcal{G}_w)"] & p_2^!\omega^\kappa \arrow[d, "p_2^*\mathrm{Ha}(\mathcal{G}_w)"] \\
p_1^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T''_{w,et}"] & p_2^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})
\end{tikzcd}$$

Proof. That $T'_{w,net} = T''_{w,net} = 0$ follows from an inspection of the proof of Lemma 3.9.22; more precisely, by the proofs of [Pil20, Lem. 7.1.1, 7.1.2], $T'_{w,net}$ is divisible by p^{k_w-2}, and $T''_{w,net}$ is divisible by p^{l_w-1}. The commutativity of the second diagram follows immediately from the fact that the Hasse invariant commutes with étale isogenies. The commutativity of the first diagram is slightly more delicate; see the proof of [Pil20, Prop. 7.4.1.1], which explains how it reduces to [Pil20, Lem. 6.3.4.1]. □

Lemma 4.2.15. *The following diagrams of locally free sheaves on $X_{K',1}^{I'_{a,b},J}$ are commutative for $l_w \geq 2$ and $k_w \geq 3$:*

$$\begin{tikzcd} p_2^*\omega^\kappa \arrow[r, "T'_w"] \arrow[d, "p_2^*\mathrm{Ha}(\mathcal{G}_w)"] & p_1^!\omega^\kappa \arrow[d, "p_1^*\mathrm{Ha}(\mathcal{G}_w)"] \\ p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T'_w"] & p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \end{tikzcd}$$

$$\begin{tikzcd} p_1^*\omega^\kappa \arrow[r, "T''_w"] \arrow[d, "p_1^*\mathrm{Ha}(\mathcal{G}_w)"] & p_2^!\omega^\kappa \arrow[d, "p_2^*\mathrm{Ha}(\mathcal{G}_w)"] \\ p_1^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "T''_w"] & p_2^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \end{tikzcd}$$

Proof. Since all sheaves are locally free and $X_{K',1}^{I'_{a,b},J}$ is Cohen–Macaulay, it is enough to check the commutativity over the dense open subscheme $(X_{K',1}^{I'_{a,b},J})^{=_w 2}$, which is Lemma 4.2.14. □

Corollary 4.2.16. *Assume that for all $v \in I'$ we have $l_v \geq 2$ and $k_v - l_v \geq C$, and that for all $v \in J$ we have $l_v \geq C$. Then the action of $T^{I_b,J}$ on $\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 2}, \omega^\kappa)$ is locally finite if $l_w \geq 2$ and $k_w \geq 3$.*

Proof. We have $H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 2}, \omega^\kappa) = \varinjlim_n H^*(X_{K,1}^{G_1,I'_{a,b},J}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{(p-1)n})$ where the transition maps (given by multiplication by $\mathrm{Ha}(\mathcal{G}_w)$) are $T^{I_b,J}$-equivariant by Lemma 4.2.15. By the inductive hypothesis, each $e(T^{I'_b,J})H^*(X_{K,1}^{G_1,I'_{a,b},J}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{(p-1)n})$ is finite-dimensional and $T^{I_b,J}$-stable, while $T^{I_b,J}$ acts locally nilpotently on $(1 - e(T^{I'_b,J}))H^*(X_{K,1}^{G_1,I'_{a,b},J}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{(p-1)n})$ (because $T^{I'_b,J}$ does). The result follows. □

Exactly as in Lemma 4.2.9, for all $l_w \geq p+1 = p-1+2$ and $k_w \geq p+2 = 3+p-1$ we obtain cohomological correspondences:

$$T'_w : p_2^*(\omega^\kappa|_{X_{K'',1}^{I'_{a,b},J,\leq_w 1}}) \to p_1^!(\omega^\kappa|_{X_{K,1}^{I'_{a,b},J,\leq_w 1}})$$

and

$$T''_w : p_1^*(\omega^\kappa|_{X_{K,1}^{I'_{a,b},J,\leq_w 1}}) \to p_2^!(\omega^\kappa|_{X_{K'',1}^{I'_{a,b},J,\leq_w 1}}).$$

We now consider the space $(X_{K'',1}^{I'_{a,b},J})^{=_w 1}$ where $\mathcal{G}_w[p]$ has p-rank 1. We have a universal quasi-polarization $\mathcal{G}_w \to \mathcal{G}_w^D$ over $X_{K''}$. Over the interior of the moduli space, the kernel of the quasi-polarization is a self dual rank p^2 group scheme which is either connected or an extension of a multiplicative by an étale group scheme. The space $(X_{K'',1}^{I'_{a,b},J})^{=_w 1}$ decomposes as the union of connected components $(X_{K'',1}^{I'_{a,b},J})^{=_w 1,00}$ and $(X_{K'',1}^{I'_{a,b},J})^{=_w 1,m\text{-}et}$ for which the kernel of the quasi-polarization doesn't contain (respectively contains) a multiplicative group (see [Pil20, Lem. 7.4.2.3]).

We now consider the space $(X_{K',1}^{I'_{a,b},J})^{=_w 1}$, which we view here only as a topological space (it has multiple natural non-reduced scheme structures defined by the

vanishing of either $p_1^*\mathrm{Ha}(\mathcal{G}_w)$ or $p_2^*\mathrm{Ha}(\mathcal{G}_w)$). We have the chain of isogenies $p_1^*\mathcal{G} \to p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ where the composite is multiplication by p. We have a decomposition of $(X_{K',1}^{I'_{a,b},J})^{=_w 1}$ as a union of connected components: $(X_{K',1}^{I'_{a,b},J})^{=_w 1,m}$, $(X_{K',1}^{I'_{a,b},J})^{=_w 1,et}$ and $(X_{K',1}^{I'_{a,b},J})^{=_w 1,00}$. Here the open and closed subspace $(X_{K',1}^{I'_{a,b},J})^{=_w 1,m}$ is the locus where the kernel of $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ is an étale group scheme; the open and closed subspace $(X_{K',1}^{I'_{a,b},J})^{=_w 1,et}$ is the locus where the kernel of $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ is a multiplicative group scheme; and the open and closed subspace $(X_{K',1}^{I'_{a,b},J})^{=_w 1,00}$ is the locus where the kernel of $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ is a bi-connected group scheme.

It follows from the definitions (see [Pil20, Lem. 7.4.2.4]) that

$$p_2((X_{K',1}^{I'_{a,b},J})^{=_w 1,00}) \subseteq (X_{K'',1}^{I'_{a,b},J})^{=_w 1,00} \tag{4.2.17}$$

and that at the level of topological spaces,

$$p_2((X_{K',1}^{I'_{a,b},J})^{=_w 1,m} \cup (X_{K',1}^{I'_{a,b},J})^{=_w 1,et}) \subseteq (X_{K'',1}^{I'_{a,b},J})^{=_w 1,m\text{-}et}. \tag{4.2.18}$$

Over $(X_{K,1}^{I'_{a,b},J})^{=_w 1}$ and $(X_{K',1}^{I'_{a,b},J})^{=_w 1}$ we can decompose the cohomological correspondences T'_w and T''_w by projecting on the various components (in other words, composing with the various idempotents associated to each of these connected components). This gives us decompositions $T'_w = T'_{w,m} + T'_{w,et} + T'_{w,00}$ and $T''_w = T''_{w,m} + T''_{w,et} + T''_{w,00}$ obtained by projecting on the multiplicative, étale and bi-connected components respectively.

Lemma 4.2.19. *The following diagrams of sheaves on $X_{K',1}^{I'_{a,b},J}$ are commutative for $l_w \geq p+1$ and $k_w \geq 2p+3$:*

$$\begin{tikzcd}
p_2^*\omega^\kappa\big|_{(X_{K'',1}^{I'_{a,b},J})^{=_w 1}} \arrow[r, "T'_{w,et}"] \arrow[d, "p_2^*\mathrm{Ha}'(\mathcal{G}_w)"] & p_1^!\omega^\kappa\big|_{(X_{K,1}^{I'_{a,b},J})^{=_w 1}} \arrow[d, "p_1^*\mathrm{Ha}'(\mathcal{G}_w)"] \\
p_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K'',1}^{I'_{a,b},J})^{=_w 1}} \arrow[r, "T'_{w,et}"] & p_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K,1}^{I'_{a,b},J})^{=_w 1}}
\end{tikzcd}$$

$$\begin{tikzcd}
p_1^*\omega^\kappa\big|_{(X_{K,1}^{I'_{a,b},J})^{=_w 1}} \arrow[r, "T''_{w,et}"] \arrow[d, "p_1^*\mathrm{Ha}'(\mathcal{G}_w)"] & p_2^!\omega^\kappa\big|_{(X_{K'',1}^{I'_{a,b},J})^{=_w 1}} \arrow[d, "p_2^*\mathrm{Ha}'(\mathcal{G}_w)"] \\
p_1^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K,1}^{I'_{a,b},J})^{=_w 1}} \arrow[r, "T''_{w,et}"] & p_2^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K'',1}^{I'_{a,b},J})^{=_w 1}}
\end{tikzcd}$$

Moreover, $T'_{w,m} = T'_{w,00} = 0$ and $T''_{w,m} = 0$. If $l_w \geq p+2$, then $T''_{w,00} = 0$.

Proof. See [Pil20, Prop. 7.4.2.1]. □

We recall that by definition we have $T_w = T'_w \circ T''_w$ as operators on the cohomology over $X_K^{I'_{a,b},J}$. It will also be useful to consider the composition $\widetilde{T}_w := T''_w \circ T'_w$ defining an operator on the cohomology over $X_{K''}^{I'_{a,b},J}$.

Lemma 4.2.20. *If $l_w \geq p+1$ and $k_w \geq 2p+3$, then we have $T_w = T'_{w,et} \circ T''_{w,et} \in \mathrm{End}(\mathrm{R}\Gamma(X_{K,1}^{I'_{a,b},J,=_w 1}, \omega^\kappa))$, and*

$$T_w \mathrm{Ha}'(\mathcal{G}_w) = \mathrm{Ha}'(\mathcal{G}_w) T_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X_{K,1}^{I'_{a,b},J,=_w 1}, \omega^\kappa), \mathrm{R}\Gamma(X_{K,1}^{I'_{a,b},J,=_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{(p^2-1)})).$$

Similarly, if $k_w \geq 2p+3$, then we have $\widetilde{T}_w = T''_{w,et} \circ T'_{w,et} \in \mathrm{End}(\mathrm{R}\Gamma(X_{K'',1}^{I'_{a,b},J,=_w 1}, \omega^\kappa))$, and

$$\widetilde{T}_w \mathrm{Ha}'(\mathcal{G}_w) = \mathrm{Ha}'(\mathcal{G}_w) \widetilde{T}_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X_{K'',1}^{I'_{a,b},J,=_w 1}, \omega^\kappa), \mathrm{R}\Gamma(X_{K'',1}^{I'_{a,b},J,=_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{(p^2-1)})).$$

Proof. We give the argument for T_w; the argument for $\widetilde{T}_w$ is essentially the same, and is left to the reader. From (4.2.17) and (4.2.18) we see that we can write T_w as the sum of the two operators $T'_{w,00} \circ T''_{w,00}$ and $(T'_{w,et} + T'_{w,m}) \circ (T''_{w,et} + T''_{w,m})$.

By Lemma 4.2.19, we have $T'_{w,m} = T'_{w,00} = 0$ and $T''_{w,m} = 0$, so that $T_w = T'_{w,et} \circ T''_{w,et}$. The commutativity with $\mathrm{Ha}'(\mathcal{G}_w)$ then follows from the commutative diagrams in Lemma 4.2.19. □

Corollary 4.2.21. *Assume that for all places $v \in I'_b$, we have $l_v \geq 2$ and $k_v \geq C$, and that for all places $v \in J$, we have $l_v \geq C$. If $l_w \geq p+1$ and $k_w \geq 2p+3$, then the action of $T^{I_b,J}$ on $\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa)$ is locally finite. Similarly, if $k_w \geq 2p+3$, then the action of $T^{I'_b,J}\widetilde{T}_w$ on $\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa)$ is also locally finite.*

Proof. Again, we give the proof for $T^{I_b,J}$, the argument for $\widetilde{T}_w T^{I',J}$ being essentially identical. We have $H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa) = \varinjlim_n H^*(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p^2-1)})$. More precisely, for all $n \geq 0$, the map $H^*(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p^2-1)}) \to H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa)$ is defined by the composition:

$$H^*(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p^2-1)}) \to H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p^2-1)})$$

$$\overset{\mathrm{Ha}'(\mathcal{G}_w)^{-n}}{\to} H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa).$$

The vector space $H^*(X_{K,1}^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa)$ is therefore an inductive limit of the images of the spaces

$$H^*(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p^2-1)})$$

under these maps. The first map is obviously $T^{I_b,J}$-equivariant, and by Lemma 4.2.20, the second map is also $T^{I_b,J}$-equivariant. The result follows from our inductive hypothesis that Theorem 4.2.12 holds for all smaller values of $\#I_b$, because by definition we have $X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1} = X_{K,1}^{G_1,\leq_{I_a \cup \{w\}} 1,\geq_{I'_b} 1,\geq_J 2}$. □

Lemma 4.2.22. *The following diagram of sheaves on $X_{K',1}^{I'_{a,b},J}$ is commutative for $l_w \geq p+1$ and $k_w \geq 2p+3$:*

$$\begin{tikzcd}
p_2^*\omega^\kappa\big|_{(X_{K'',1}^{I'_{a,b},J})\leq_w 1} \arrow[r, "T'_w"] \arrow[d, "p_2^*\mathrm{Ha}'(\mathcal{G}_w)"] & p_1^!\omega^\kappa\big|_{(X_{K,1}^{I'_{a,b},J})\leq_w 1} \arrow[d, "p_1^*\mathrm{Ha}'(\mathcal{G}_w)"] \\
p_2^*(\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K'',1}^{I'_{a,b},J})\leq_w 1} \arrow[r, "T'_w"] & p_1^!(\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K,1}^{I'_{a,b},J})\leq_w 1}
\end{tikzcd}$$

If $l_w \geq p+2$ and $k_w \geq 2p+3$, the following diagram is commutative:

$$\begin{tikzcd}
p_1^*\omega^\kappa\big|_{(X_{K',1}^{I'_{a,b},J})\leq_w 1} \arrow[r, "T''_w"] \arrow[d, "p_1^*\mathrm{Ha}'(\mathcal{G}_w)"] & p_2^!\omega^\kappa\big|_{(X_{K'',1}^{I'_{a,b},J})\leq_w 1} \arrow[d, "p_2^*\mathrm{Ha}'(\mathcal{G}_w)"] \\
p_1^*(\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K',1}^{I'_{a,b},J})\leq_w 1} \arrow[r, "T''_w"] & p_2^!(\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})\big|_{(X_{K'',1}^{I'_{a,b},J})\leq_w 1}
\end{tikzcd}$$

Proof. It is enough to check the commutativity over a dense open subscheme of the support of these Cohen–Macaulay sheaves, and this follows from Lemma 4.2.19. □

Since $p_1^*\mathrm{Ha}'(\mathcal{G}_w)$ is not a zero divisor on $X_{K',1}|_{(X_{K,1}^{I'_{a,b},J})\leq_w 1}$, it follows exactly as in the proof of Lemma 4.2.9 that there is for all $l_w \geq p^2+p$ and $k_w \geq p^2+2p+2$ a cohomological correspondence:

$$T'_w : p_2^*(\omega^\kappa|_{(X_{K'',1}^{I'_{a,b},J})=_w 0}) \to p_1^!(\omega^\kappa|_{(X_{K,1}^{I'_{a,b},J})=_w 0})$$

and similarly for all $l_w \geq p^2+p+1$ and $k_w \geq p^2+2p+2$ a cohomological correspondence:

$$T''_w : p_1^*(\omega^\kappa|_{(X_{K,1}^{I'_{a,b},J})=_w 0}) \to p_2^!(\omega^\kappa|_{(X_{K'',1}^{I'_{a,b},J})=_w 0}).$$

Lemma 4.2.23. *There is a universal constant C' which depends only on F and p but not on the tame level such that*

$$T'_w : p_2^*(\omega^\kappa|_{(X_{K'',1}^{I'_{a,b},J})=_w 0}) \to p_1^!(\omega^\kappa|_{(X_{K,1}^{I'_{a,b},J})=_w 0})$$

is zero for all $k_w - l_w \geq C'$ and all $l_w \geq p^2+p$.

Proof. See [Pil20, Prop. 7.4.2.2]. □

We now increase our constant C if necessary, so that $C \geq C'$, where C' is as in Lemma 4.2.23.

Lemma 4.2.24. *Assume that for all places $v \in I'_b$, we have $l_v \geq 2$ and $k_v - l_v \geq C$, that $l_v \geq C$ for all $v \in J$, and that $l_w \geq p+2$ and $k_w - l_w \geq C$. Then the map $e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \to e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)$ is a quasi-isomorphism. In particular, $e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)$ is a perfect complex.*

Proof. Consider the following diagram of exact triangles:

$$\begin{tikzcd}
\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \arrow[r, "T'_w"] \arrow[d, "\mathrm{Ha}'(\mathcal{G}_w)"] & \mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \arrow[d, "\mathrm{Ha}'(\mathcal{G}_w)"] \\
\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[r, "T'_w"] \arrow[d] & \mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[d] \\
\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,=_w 0},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[r, "T'_w"] & \mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 0},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})
\end{tikzcd}$$

By Lemma 4.2.23, the rightmost operator T'_w acts by zero. We have the ordinary projectors $e(T^{I_b,J})$ on $\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$, and the ordinary projectors $e(\widetilde{T}_w T^{I'_b,J})$ on $\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)$ and $\mathrm{R}\Gamma(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$. It follows from the defining properties of the ordinary projectors that after applying them, the left two vertical arrows T'_w are quasi-isomorphisms.

By Lemma 4.2.22 the projectors commute with multiplication by $\mathrm{Ha}'(\mathcal{G}_w)$. It follows from a short diagram chase that the map

$$e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)\overset{\mathrm{Ha}'(\mathcal{G}_w)}{\to} e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$$

is a quasi-isomorphism. The claimed quasi-isomorphism now follows by taking an inductive limit as in the proof of Corollary 4.2.21. By our inductive hypothesis, $e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)$ is a perfect complex, so we are done. □

Lemma 4.2.25. *Assume that for all places $v\in I$, we have $l_v\geq 2$ and $k_v-l_v\geq C$, that $l_v\geq C$ for all $v\in J$, and that $l_w=p+1$. Then the ordinary cohomology $e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)$ is a perfect complex, and the map*

$$e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)\to e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)$$

is an isomorphism.

Proof. The map

$$\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)\overset{\mathrm{Ha}'(\mathcal{G}_w)}{\to}\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$$

is a quasi-isomorphism, which commutes with the projector $e(T^{I_b,J})$ by Lemma 4.2.20. It follows from Lemma 4.2.24 that $e(T^{I_b,J})\mathrm{R}\Gamma(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa)$ is perfect.

We now prove the claimed isomorphism on degree 0 cohomology. Since $X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1}$ is Cohen–Macaulay, and $X_{K,1}^{G_1,I'_{a,b},J,=_w 1}$ is an open dense subscheme, we have injections

$$H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\mathcal{F})\hookrightarrow H^0(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\mathcal{F})$$

for any locally free sheaf $\mathcal{F}$, so it is enough to prove surjectivity. In order to do this, it is enough to prove that for all $n\geq 0$, the map

$$H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)\overset{(\mathrm{Ha}'(\mathcal{G}_w))^n}{\to}H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p^2-1)})$$

(which commutes with $e(T_w)$ by Lemma 4.2.20 and the injectivity of the restrictions $H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\mathcal{F}) \hookrightarrow H^0(X_{K,1}^{G_1,I'_{a,b},J,=_w 1},\mathcal{F})$ discussed above) induces a surjection:

$$e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \overset{(\mathrm{Ha}'(\mathcal{G}_w))^n}{\to} e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p^2-1)}).$$

In fact, by Lemma 4.2.24, it suffices to prove the surjectivity for $n = 1$. We consider the following diagram:

$$\begin{tikzcd}
H^0(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \arrow[r, "T'_w"] \arrow[d, "\mathrm{Ha}'(\mathcal{G}_w)"] & H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa) \arrow[d, "\mathrm{Ha}'(\mathcal{G}_w)"] \\
H^0(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[r, "T'_w"] \arrow[d] & H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[d] \\
H^0(X_{K'',1}^{G_1,I'_{a,b},J,=_w 0},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \arrow[r, "T'_w"] & H^0(X_{K,1}^{G_1,I'_{a,b},J,=_w 0},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})
\end{tikzcd}$$

Let $f \in e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$. As in the proof of Lemma 4.2.24, T'_w induces a bijection

$$e(\widetilde{T}_w T^{I'_b,J})H^0(X_{K'',1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}) \to e(T^{I_b,J})H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1}).$$

In particular, $f = T'_w g$ for some $g \in H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$ and therefore, since the rightmost operator T'_w acts by zero by Lemma 4.2.23, f has trivial image in $H^0(X_{K,1}^{G_1,I'_{a,b},J,=_w 0},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{p^2-1})$. It follows that f comes from a class $\tilde{f} \in H^0(X_{K,1}^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa)$. Replacing $\tilde{f}$ by $e(T^{I_b,J})\tilde{f}$ we deduce the required surjectivity. □

Corollary 4.2.26. *Assume that for all places $v \in I_b$, we have $l_v \geq 2$ and $k_v - l_v \geq C$, and that for all places $v \in J$, we have $l_v \geq C$. Then $T^{I_b,J}$ acts locally finitely on $\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa)$ and $e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa)$ is a perfect complex.*

We have an isomorphism:

$$e(T^{I_b,J})H^0(X_1^{G_1,\leq_{I_a} 1},\omega^\kappa) \xrightarrow{\sim} e(T^{I_b,J})H^0(X_1^{G_1,I_{a,b},J},\omega^\kappa)$$

and an injection:

$$e(T^{I_b,J})H^1(X_1^{G_1,\leq_{I_a} 1},\omega^\kappa) \hookrightarrow e(T^{I_b,J})H^1(X_1^{G_1,I_{a,b},J},\omega^\kappa).$$

If furthermore $l_w \geq 3$, then the map

$$e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J},\omega^\kappa) \to e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa)$$

is a quasi-isomorphism.

Proof. We begin by showing that $T^{I_b,J}$ acts locally finitely on both $\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 2},\omega^\kappa)$ and $\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa) = \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)$. For $\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 2},\omega^\kappa)$, this is Corollary 4.2.16.

Our argument for $\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)$ is slightly more involved. We have an exact triangle

$$\begin{tikzcd} \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa) \arrow[r] & \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 2},\omega^\kappa) \arrow[d] \\ & \mathrm{R}\Gamma_{X^{G_1,I'_{a,b},J,=_w 1}}(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)[+1], \end{tikzcd}$$

so it is enough to prove that the action of $T^{I_b,J}$ on $\mathrm{R}\Gamma_{X_1^{G_1,I'_{a,b},J,=_w 1}}(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)[+1]$ is locally finite. We have

$$\begin{tikzcd} \mathrm{R}\Gamma_{X_1^{G_1,I'_{a,b},J,=_w 1}}(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)[+1] \arrow[d, equal] \\ \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\varinjlim_n \omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)}) \end{tikzcd} \tag{4.2.27}$$

so it is enough to prove that the action of $T^{I_b,J}$ on each $\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)})$ is locally finite. In the case $n=1$, this is Corollary 4.2.21, and the general case follows by induction by taking the cohomology of the short exact sequence of sheaves

$$\begin{aligned} 0\to\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{(n-1)(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^{n-1})} &\xrightarrow{\times\mathrm{Ha}(\mathcal{G}_w)} \omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)} \\ &\to \omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w))}\to 0 \end{aligned} \tag{4.2.28}$$

Consider now the following diagram of exact triangles:

$$\begin{tikzcd} \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J},\omega^\kappa) \arrow[r] \arrow[d] & \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa) \arrow[d] \\ \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 2},\omega^\kappa) \arrow[r] \arrow[d] & \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 2},\omega^\kappa) \arrow[d] \\ \mathrm{R}\Gamma_{X_1^{G_1,I'_{a,b},J,\leq_w 1}}(X_1^{G_1,I'_{a,b},J},\omega^\kappa)[+1] \arrow[r] & \mathrm{R}\Gamma_{X_1^{G_1,I'_{a,b},J,=_w 1}}(X_1^{G_1,I'_{a,b},J,\geq_w 1},\omega^\kappa)[+1] \end{tikzcd}$$

We have already seen that $T^{I_b,J}$ acts locally finitely on all but the last term of the first row, so it acts locally finitely on every term in the diagram. The middle vertical arrow is the identity map. In order to show that the left vertical arrow is a quasi-isomorphism after applying $e(T^{I_b,J})$, it is therefore enough to prove it for the right vertical arrow. By (4.2.27) and the similar expression

$$\mathrm{R}\Gamma_{X_1^{G_1,I'_{a,b},J,\leq_w 1}}(X_1^{G_1,I'_{a,b},J},\omega^\kappa)[+1]=\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J},\varinjlim_n\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)}),$$

it suffices to show that for each $n\geq 1$, the map

$$\begin{array}{c} e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)}) \\ \downarrow \\ e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{n(p-1)}|_{V(\mathrm{Ha}(\mathcal{G}_w)^n)}) \end{array}$$

is a quasi-isomorphism. To see this, note that the case $n = 1$ is Lemma 4.2.24, and the general case follows by induction on n, by taking the cohomology of the exact sequence of sheaves (4.2.28). The remaining claims follow from the quasi-isomorphism just proved and the inductive hypothesis. □

Corollary 4.2.29. *Assume that for all places $v \in I$, we have $k_v - l_v \geq C$ and $l_v \geq 2$, that for all places $v \in J$, we have $l_v \geq C$. Then the ordinary cohomology*

$$e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$$

is represented by a perfect complex. Moreover we have an isomorphism:

$$e(T^{I_b,J})H^0(X_1^{G_1,\leq_{I_a} 1}, \omega^\kappa) \xrightarrow{\sim} e(T^{I_b,J})H^0(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$$

and an injection:

$$e(T^{I_b,J})H^1(X_1^{G_1,\leq_{I_a} 1}, \omega^\kappa) \hookrightarrow e(T^{I_b,J})H^1(X_1^{G_1,I_{a,b},J}, \omega^\kappa).$$

Proof. To see that

$$e(T^{I_b,J})\mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$$

is represented by a perfect complex, we consider the exact triangle:

$$\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1}, \omega^\kappa) \overset{\mathrm{Ha}(\mathcal{G}_w)}{\to} \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\geq_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})$$

$$\to \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,=_w 1}, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \overset{+1}{\to}$$

Applying the projector $e(T^{I_b,J})$ everywhere (which commutes with the various maps by Lemma 4.2.15) we deduce this from Corollary 4.2.26 and Lemma 4.2.25. By our inductive hypothesis, in order to prove the claims about the morphisms

$$e(T^{I_b,J})H^i(X_1^{G_1,\leq_{I_a} 1}, \omega^\kappa) \xrightarrow{\sim} e(T^{I_b,J})H^i(X_1^{G_1,I_{a,b},J}, \omega^\kappa),$$

it is enough to prove the corresponding statements for the morphisms

$$e(T^{I_b,J})H^i(X_1^{G_1,I'_{a,b},J}, \omega^\kappa) \to e(T^{I_b,J})H^i(X_1^{G_1,I_{a,b},J}, \omega^\kappa).$$

Firstly, the natural restriction map

$$H^0(X_1^{G_1,I'_{a,b},J}, \omega^\kappa) \to H^0(X_1^{G_1,I_{a,b},J}, \omega^\kappa)$$

is an isomorphism, because $X_1^{G_1,I'_{a,b},J}$ is Cohen–Macaulay, and the complement of $X_1^{G_1,I_{a,b},J}$ is of codimension at least 2. It remains to prove the injectivity of the map of H^1s.

We have a commutative diagram of exact triangles:

$$\begin{array}{ccc}
\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J},\omega^\kappa) & \longrightarrow & \mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa) \\
\downarrow & & \downarrow \\
\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{(p-1)}) & \longrightarrow & \mathrm{R}\Gamma(X_1^{G_1,I_{a,b},J},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{(p-1)}) \\
\downarrow & & \downarrow \\
\mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,\leq_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{(p-1)}) & \longrightarrow & \mathrm{R}\Gamma(X_1^{G_1,I'_{a,b},J,=_w 1},\omega^\kappa\otimes\det\omega_{\mathcal{G}_w}^{(p-1)})
\end{array}$$

The injectivity of $e(T^{I_b,J})H^1(X_1^{G_1,I'_{a,b},J},\omega^\kappa)\to e(T^{I_b,J})H^1(X_1^{G_1,I_{a,b},J},\omega^\kappa)$ therefore follows from a short diagram chase, using the quasi-isomorphisms provided by Corollary 4.2.26 and the isomorphism on H^0s of Lemma 4.2.25. □

Proof of Theorem 4.2.12. This is immediate from Corollaries 4.2.26 and 4.2.29. □

4.2.30. *A Cousin complex computing* $\mathrm{R}\Gamma(X_1^{G_1,I},\omega^\kappa(-D))$. Our goal in this section is to provide an explicit Hecke stable complex computing $\mathrm{R}\Gamma(X_1^{G_1,I},\omega^\kappa(-D))$. This complex will be used to complete the proof of Theorem 4.2.1, and will also be used in §4.6 to compare the cohomology at spherical and Klingen levels (by considering the corresponding complex at Klingen level, and the natural map between these complexes). This complex is the Cousin complex associated to $X_1^{G_1,I}$ and the stratification given by the p-rank (see §3.9.5). This section is very similar to §3.9.10 where we introduced the Cousin complex over the full Shimura variety associated with the Ekedahl–Oort stratification. The case we consider here is, however, much simpler because we have canonical global equations provided by the partial Hasse invariants for our stratification. We has thus decided, despite redundancy, to give a complete and explicit construction of the Cousin complex in this case.

Let S be a smooth scheme over a field k and let $\mathcal{L}$ be an invertible sheaf on S. We assume that $\mathcal{L}=\otimes_{i=1}^d\mathcal{L}_i$ and that we have non-vanishing sections $s_i\in H^0(S,\mathcal{L}_i)$. We let $D_i=V(s_i)$, an effective Cartier divisor on S. Set $s=\prod_{i=1}^d s_i$. Set $D=V(s)=\cup_i D_i$. We assume that $D=\cup_i D_i$ is a strict normal crossing divisor on S.

For all n, consider the following exact complex of coherent sheaves on S:

$$0\to\mathcal{O}_S\xrightarrow{s^n}\mathcal{L}^n\to\bigoplus_{i=1}^d\mathcal{L}^n/s_i^n\to\bigoplus_{1\leq i<j\leq d}\mathcal{L}^n/(s_i^n,s_j^n)\to\cdots\to\mathcal{L}^n/(s_1^n,\cdots,s_d^n)\to 0$$

This is a complex of length $d+2$. For all $0\leq k\leq d$, the object placed in degree $k+1$ is $\bigoplus_{1\leq i_1<\cdots<i_k\leq d}\mathcal{L}^n/(s_{i_1}^n,\cdots,s_{i_k}^n)$ (when we write $\mathcal{L}^n/(s_{i_1}^n,\cdots,s_{i_k}^n)$, we mean $\mathcal{L}^n/\mathcal{L}^n(s_{i_1}^n\mathcal{L}_{i_1}^{-n},\cdots,s_{i_k}^n\mathcal{L}_{i_k}^{-n})$). The differential

$$\bigoplus_{1\leq i_1<\cdots<i_k\leq d}\mathcal{L}^n/(s_{i_1}^n,\cdots,s_{i_k}^n)\to\bigoplus_{1\leq i_1<\cdots<i_{k+1}\leq d}\mathcal{L}^n/(s_{i_1}^n,\cdots,s_{i_k}^n)$$

takes a section $(f_{i_1,\cdots,i_k})_{1\leq i_1<\cdots<i_k\leq d}$ to the section

$$(\sum(-1)^{i_j}\overline{f_{i_1,\cdots,\widehat{i_j},\cdots,i_{k+1}}})_{1\leq i_1<\cdots<i_{k+1}\leq d}$$

where $\overline{f_{i_1,\cdots,\widehat{i_j},\cdots,i_{k+1}}}$ is the class modulo s_j^n of $f_{i_1,\cdots,\widehat{i_j},\cdots,i_{k+1}}$.

The following diagram is commutative:

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{O}_S \arrow[r, "s^{n+1}"] & \mathcal{L}^{n+1} \arrow[r] & \bigoplus_{i=1}^d \mathcal{L}^{n+1}/s_i^{n+1} \arrow[r] & \cdots \\
0 \arrow[r] & \mathcal{O}_S \arrow[r, "s^n"] \arrow[u, "\mathrm{id}"] & \mathcal{L}^n \arrow[r] \arrow[u, "s"] & \bigoplus_{i=1}^d \mathcal{L}^n/s_i^n \arrow[r] \arrow[u, "s"] & \cdots
\end{tikzcd}$$

Passing to the limit over n, we get the following exact complex:

$$0 \to \mathcal{O}_S \to \varinjlim_n \mathcal{L}^n \to \varinjlim_n \bigoplus_{i=1}^d \mathcal{L}^n/s_i^n \to$$

$$\varinjlim_n \bigoplus_{1\le i<j\le d} \mathcal{L}^n/(s_i^n, s_j^n) \to \cdots \to \varinjlim_n \mathcal{L}^n/(s_1^n, \cdots, s_d^n) \to 0$$

where in all the direct limits, the transition maps are given by multiplication by powers of s.

Lemma 4.2.31. *Let $1 \le i_1 < \cdots < i_k \le d$. Set $\mathcal{L}' = \otimes_{j=1}^k \mathcal{L}_{i_j}$, $s' = \prod_{j=1}^k s_{i_j}$, $D' = V(s(s')^{-1})$. There is a canonical isomorphism*

$$\varinjlim_{\times s^n} \mathcal{L}^n/(s_{i_1}^n, \cdots, s_{i_k}^n) \simeq \varinjlim_{\times (s')^n} ((\mathcal{L}')^n/(s_{i_1}^n, \cdots, s_{i_k}^n))|_{S\setminus D'}$$

Proof. Easy and left to the reader. □

Remark 4.2.32. The complex

$$0 \to \varinjlim_n \mathcal{L}^n \to \varinjlim_n \bigoplus_{i=1}^d \mathcal{L}^n/s_i^n \to$$

$$\varinjlim_n \bigoplus_{1\le i<j\le d} \mathcal{L}^n/(s_i^n, s_j^n) \to \cdots \to \varinjlim_n \mathcal{L}^n/(s_1^n, \cdots, s_d^n) \to 0$$

is just the Cousin complex of $\mathcal{O}_S$ associated with the stratification given by the divisors D_i.

We now work over $S = X_1^{G_1,I}$. We take $\mathcal{L} = \otimes_{w\in I}(\det \mathcal{G}_w)^{p-1}$, $\mathcal{L}_w = (\det \mathcal{G}_w)^{p-1}$ and $s_w = \mathrm{Ha}(\mathcal{G}_w)$, and we consider the complex $K^0 \to K^1 \cdots \to K^d$ obtained by applying H^0 to the complex

$$\varinjlim_n \mathcal{L}^n \to \varinjlim_n \bigoplus_{i=1}^d \mathcal{L}^n/s_i^n \to \varinjlim_n \bigoplus_{1\le i<j\le d} \mathcal{L}^n/(s_i^n, s_j^n) \to \cdots \to \varinjlim_n \mathcal{L}^n/(s_1^n, \cdots, s_d^n)$$

tensored with $\omega^\kappa(-D)$. (So in the above notation, the indices i will correspond to the different places $v \in I$, the s_i will correspond to Hasse invariants, and the assumption that the divisor $V(s)$ has strict normal crossings is an easy consequence of the Serre–Tate theorem and the product structure on the p-divisible group $\mathcal{G}$.)

It follows from Lemma 4.2.31 that K^k equals

$$\bigoplus_{J\subset I, \#J=k} \varinjlim_{\times \prod_{w\in J} \mathrm{Ha}(\mathcal{G}_w)} H^0\big(X_1^{G_1,I,\ge_{J^c}2}, \omega^\kappa(-D) \otimes \bigotimes_{w\in J} (\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J} (\mathrm{Ha}(\mathcal{G}_w)^n))\big).$$

Proposition 4.2.33. *The complex $K^\bullet$ computes $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$.*

Proof. The argument is the same as in the proof of Proposition 3.9.11. It suffices to show that each of the sheaves

$$\omega^\kappa(-D) \otimes \bigotimes_{w\in J} (\det \mathcal{G}_w)^{n(p-1)} / (\sum_{w\in J} (\mathrm{Ha}(\mathcal{G}_w)^n))$$

when restricted to $X_1^{G_1,I,\geq_{J^c}2}$ and then pushed forward to $X_1^{G_1,I}$ is acyclic on $X_1^{G_1,I}$. Since the inclusion $X_1^{G_1,I,\geq_{J^c}2} \hookrightarrow X_1^{G_1,I}$ is affine, it suffices to show that the restriction of this sheaf to $X_1^{G_1,I,\geq_{J^c}2}$ is acyclic. By [Lan17, Thm. 8.6], this sheaf is acyclic relative to the minimal compactification and its support in the minimal compactification is the locally closed subscheme given by the set of equations:

- $\mathrm{Ha}(\mathcal{G}_w)^n = 0$ for $w \in J$,
- $\mathrm{Ha}'(\mathcal{G}_w) \neq 0$ for $w \in J$,
- $\mathrm{Ha}(\mathcal{G}_w) \neq 0$ for $w \in J^c$,

which is affine. □

Remark 4.2.34. Using Proposition 4.2.33, one can show that if we have $l_w \geq 2$ and $k_w \geq 2p+3$ for $w \in I$, and $l_w \geq 3$ for $w \in I^c$, then the individual Hecke operators T_w for $w \in I$ and $T_{w,1}$ for $w \in I^c$ act locally finitely on $R\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ (by showing that they act locally finitely on each term of the complex $K^\bullet$). We leave the details to the interested reader.

We can finally complete the proof of Theorem 4.2.1.

Proof of Theorem 4.2.1. Everything is immediate from Theorem 4.2.12 (taking $I_a = \emptyset$ and $J = I^c$), except for the claim that $e(T^I)\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ has amplitude $[0, \#I]$, which follows from Proposition 4.2.33. □

4.2.35. *Commutativity over the ordinary locus.* While we do not prove the commutativity of the correspondences $T_{w,1}$ and T_w, we do prove it over the ordinary locus at w, where all of the correspondences are finite flat over the interior. We will need this result at the places $w \in I^c$, because we need to make use of both of these Hecke operators in this case (because the Hecke operator $U_{w,2}$ at Klingen level which corresponds to T_w is needed for those parts of the control theorem which take place at the level of the sheaf, but we need to use $T_{w,1}$ to prove the finiteness of cohomology).

Lemma 4.2.36. *Suppose that $w \in I^c$, and that $l_w \geq 2$. Then on $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ we have $T_{w,1} \circ T_w = T_w \circ T_{w,1}$.*

Proof. We can easily compose cohomological correspondences when the projections are finite flat. In particular we may form the compositions $T_w \circ T_{w,1}$ and $T_{w,1} \circ T_w$ over the interior, and it is easy to see that the compositions give the same cohomological correspondence.

In order to check that they commute on $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$ we use a similar trick to the one that we used to prove Proposition 3.9.15: recall the complex $K^\bullet$ of Proposition 4.2.33 which computes $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^\kappa(-D))$. We may form another complex $K'^\bullet$ by applying the same construction to the interior $Y_1^{G_1,I} \subset X_1^{G_1,I}$. As we have explained above, the Hecke operators T_w and $T_{w,1}$ commute on each term

$$H^0\big(Y_1^{G_1,I,\geq_{J^c}2}, \omega^\kappa(-D) \otimes \bigotimes_{w\in J} (\det \mathcal{G}_w)^{n(p-1)} / (\sum_{w\in J} (\mathrm{Ha}(\mathcal{G}_w)^n))\big)$$

in the definition of $K'^{\bullet}$, and hence on the subcomplex $K^{\bullet}$ of $K'^{\bullet}$ and thus on $\mathrm{R}\Gamma(X_1^{G_1,I}, \omega^{\kappa}(-D))$. □

We end this section by proving the following technical result, whose formulation relies on Lemma 4.2.36. We will make use of it in §4.6, in order to compare the complex of Proposition 4.2.33 to the analogous complex at Klingen level. Fix a subset $J \subset I$; we now consider the space $X_{K,1}^{I,=_J 1,=_{J^c} 2} = X_{K,1}^{=_J 1,=_{J^c} 2}$.

Lemma 4.2.37. *There is a universal constant C depending only on p and F but not on the tame level K^p such that if $l_v \geq 2$, $k_v - l_v \geq C$ for all $v \in I$, $l_v \geq C$ for all $v \in I^c$ and $l_v \geq p+1$ for all $v \in J$, then $\mathrm{R}\Gamma(X_{K,1}^{I,=_J 1,=_{J^c} 2}, \omega^{\kappa}(-D))$ carries a locally finite action of $\widetilde{T}^I = \prod_{w|p} T_w \prod_{w \in I^c} T_{w,1}$.*

Proof. Note that by Lemma 4.2.36, all of the Hecke operators in the definition of $\widetilde{T}^I$ commute. We begin by showing that the action of $\widetilde{T}^I$ on $\mathrm{R}\Gamma(X_{K,1}^I, \omega^{\kappa}(-D))$ is locally finite. To this end, note that by Theorem 4.2.12, the action of T^I on $\mathrm{R}\Gamma(X_{K,1}^I, \omega^{\kappa}(-D))$ is locally finite, and $e(T^I)\mathrm{R}\Gamma(X_{K,1}^I, \omega^{\kappa}(-D))$ is a perfect complex if $l_v \geq 2$, $k_v - l_v \geq C$ for all $v \in I$, and $l_v \geq C$ for all $v \in I^c$. Since $\widetilde{T}^I = T^I \prod_{w \in I^c} T_w$, it follows that the action of $\widetilde{T}^I$ is also locally finite (as it acts locally nilpotently on $(1 - e(T^I))\mathrm{R}\Gamma(X_{K,1}^I, \omega^{\kappa}(-D))$).

Taking the exact triangles induced by

$$\omega^{\kappa} \overset{\mathrm{Ha}(\mathcal{G}_w)}{\to} \omega^{\kappa} \otimes \det \omega_{\mathcal{G}_w}^{p-1} \to \omega^{\kappa} \otimes \det \omega_{\mathcal{G}_w}^{p-1} / \mathrm{Ha}(\mathcal{G}_w)$$

for all $w \in J$, and using Lemma 4.2.14, we deduce that $\widetilde{T}^I$ is locally finite on $\mathrm{R}\Gamma(X_{K,1}^{I,=_J 1}, \omega^{\kappa}(-D))$ for all weights $\kappa = (k_v, l_v)$ with $l_v \geq 2$, $k_v - l_v \geq C$ for all $v \in I$, $l_v \geq C$ for all $v \in I^c$ and $l_v \geq p+1$ if $v \in J$. Passing to the limit over multiplication by $\mathrm{Ha}(\mathcal{G}_w)$ for $w \in I \setminus J$, we deduce that $\widetilde{T}^I$ is locally finite on $\mathrm{R}\Gamma(X_{K,1}^{I,=_J 1,=_{J^c} 2}, \omega^{\kappa}(-D))$, as required. □

4.3. **Formal geometry.** In this section we continue to assume that $K = K^p K_p$, $K_p = \prod_{v|p} K_v$ with $K_v \in \{\mathrm{GSp}_4(\mathcal{O}_{F_v}), \mathrm{Par}(v)\}$. Our goal in this section is to define the Igusa tower at Klingen level, and the p-adic sheaves whose cohomology defines our spaces of p-adic automorphic forms.

4.3.1. *Completion of X.* We adopt the convention that if Z is a scheme over $\mathrm{Spec}\, \mathbf{Z}_{(p)}$, then we write Z_n for $Z \otimes_{\mathbf{Z}_{(p)}} \mathbf{Z}/p^n\mathbf{Z}$, and $\mathfrak{Z} := \varinjlim_n Z_n$ for the formal p-adic completion of Z, which is by definition a p-adic formal scheme. In particular, we let $\mathfrak{X}_K$ be the formal p-adic completion of X_K, and we write $\mathfrak{X}_K^{\geq 2} \hookrightarrow \mathfrak{X}_K^{\geq 1} \hookrightarrow \mathfrak{X}_K$ for the open formal subschemes corresponding to $X_{K,1}^{\geq 1}$ and $X_{K,1}^{\geq 2}$. We write $\mathfrak{Y}_K$ for the complement of the boundary of $\mathfrak{X}_K$, with special fibre $Y_{K,1}$. We write $\mathfrak{Y}_K^{\geq 2}$ for the ordinary locus on the interior, and so on.

4.3.2. *Deep Klingen level structure.* For all $m \geq 1$ we consider the formal scheme $\mathfrak{X}_{K,\mathrm{Kli}}^{\geq 1}(p^m) \to \mathfrak{X}_K^{\geq 1}$ which parametrizes a subgroup $H_m \subset \mathcal{G}[p^m]$ which is locally for the étale topology isomorphic to $\mu_{p^m} \otimes \mathcal{O}_F$; equivalently, $H_m = \prod_{w|p} H_{m,w}$ where for each $w|p$, $H_{m,w} \subset \mathcal{G}_w[p^m]$ is isomorphic to μ_{p^m}.

Proposition 4.3.3. *The morphism $\mathfrak{X}_{K,\mathrm{Kli}}^{\geq 1}(p^m) \to \mathfrak{X}_K^{\geq 1}$ is affine and étale. Its fibre $\mathfrak{X}_{K,\mathrm{Kli}}^{\geq 2}(p^m)$ over $\mathfrak{X}_K^{\geq 2}$ is finite étale.*

Proof. This can be proved in exactly the same way as [Pil20, Lem. 9.1.1.1]. □

We denote by $\mathfrak{X}^{\geq 1}_{K,\mathrm{Kli}}(p^\infty) = \varprojlim_m \mathfrak{X}^{\geq 1}_{K,\mathrm{Kli}}(p^m)$ the p-adic formal scheme obtained by taking the inverse limit (in the category of p-adic formal schemes). It exists because the transition morphisms are affine (see for example [Far08, Prop. D.4.1], or [Sta13, Tag 01YT] for the corresponding statement for schemes, from which this follows easily). Over $\mathfrak{X}^{\geq 1}_{K,\mathrm{Kli}}(p^\infty)$ we have for all places $v|p$ a Barsotti–Tate group of height one and dimension one $H_{\infty,v} \hookrightarrow \mathcal{G}_v$.

4.3.4. *Igusa towers.* We fix a partition $\{v|p\} = I \coprod I^c$, and we let $X^{\geq_I 1,\mathrm{Par}\text{-}m\text{-}et,\geq_{I^c} 2}_{K,1}$ be the open subscheme of $X^{\geq_I 1,\geq_{I^c} 2}_{K,1}$ where for each place $v \in I$ with $K_v = \mathrm{Par}(\mathcal{O}_{F_v})$, the kernel of the quasi-polarization $\lambda : \mathcal{G}_v \to \mathcal{G}_v^D$ contains a multiplicative group (so away from the boundary, this kernel is an extension of an étale group of rank p by a multiplicative group of rank p). We then let $\mathfrak{X}^{\geq_I 1,\mathrm{Par}\text{-}m\text{-}et,\geq_{I^c} 2}_{K}$ be the corresponding open of $\mathfrak{X}_K$, and in order to save some notation, we will for the moment set $\mathfrak{X}^I_K := \mathfrak{X}^{\geq_I 1,\mathrm{Par}\text{-}m\text{-}et,\geq_{I^c} 2}_{K}$. The fibre of $\mathfrak{X}^{\geq 1}_{K,\mathrm{Kli}}(p^m)$ over $\mathfrak{X}^I_K$ is denoted by $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. Over $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ we have for all places $v \in I$ a Barsotti–Tate group of height one and dimension one $H_{\infty,v} \hookrightarrow \mathcal{G}_v$. Observe that for all $v \in I^c$, we have a rank 2 multiplicative Barsotti–Tate group $\mathcal{G}_v^m \hookrightarrow \mathcal{G}_v$ and that $H_{\infty,v} \hookrightarrow \mathcal{G}_v^m$ is a rank one sub-Barsotti–Tate group.

If $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ for all $v|p$, this gives us a convenient alternative description of $\mathfrak{X}^I_{K,\mathrm{Kli}}(p)$. Set

$$K_p(I) = \prod_{v\in I} \mathrm{Kli}(v) \prod_{v\in I^c} \mathrm{Iw}(v).$$

Then we have $\mathfrak{X}^I_{K,\mathrm{Kli}}(p) = \mathfrak{X}^I_{K_p(I)K^p}$, where the superscript I refers to the fact that for each $v \in I$, the Klingen level structure H_v is multiplicative, and at each $v \in I^c$, we have extended the given multiplicative Klingen level structure to the canonical (ordinary) Iwahori level structure.

We denote by $\mathfrak{IG}^I \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ the profinite-étale torsor of trivializations:

$$\psi_v : \mathbf{Z}_p \simeq T_p(H^D_{\infty,v}),\ v|p;\ \phi_v : \mathbf{Z}_p \simeq T_p((\mathcal{G}_v^m/H_{\infty,v})^D),\ v \in I^c.$$

The upper script D stands for the dual of these Barsotti–Tate group schemes and T_p stands for the Tate module which here is a pro-étale sheaf. For all $v \in I$, there is an action of $\lambda \in \mathbf{Z}_p^\times$ on ψ_v, mapping ψ_v to $\psi_v \circ \lambda$. For all $v \in I^c$, there is an action of $(\lambda, \mu) \in (\mathbf{Z}_p^\times)^2$ on (ψ_v, ϕ_v), mapping (ψ_v, ϕ_v) to $(\psi_v \circ \lambda, \phi_v \circ \mu)$.

The Galois group of the torsor $\mathfrak{IG}^I \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ is

$$T_I := \prod_{v\in I} \mathcal{O}_{F_v}^\times \prod_{v\in I^c} (\mathcal{O}_{F_v}^\times)^2 \simeq \prod_{v\in I} \mathbf{Z}_p^\times \prod_{v\in I^c} (\mathbf{Z}_p^\times)^2.$$

4.3.5. *Sheaves of p-adic modular forms.* Let $\widetilde{\Lambda}_{1,v} = \mathbf{Z}_p[[\mathcal{O}_{F_v}^\times]] \simeq \mathbf{Z}_p[[\mathbf{Z}_p^\times]]$ and $\widetilde{\Lambda}_{2,v} = \mathbf{Z}_p[[(\mathcal{O}_{F_v}^\times)^2]] \simeq \mathbf{Z}_p[[(\mathbf{Z}_p^\times)^2]]$ be the one and two variable Iwasawa algebras. Let

$$\widetilde{\Lambda}_I = \hat{\otimes}_{v\in I} \widetilde{\Lambda}_{1,v} \hat{\otimes}_{v\in I^c} \widetilde{\Lambda}_{2,v} = \mathbf{Z}_p[[T_I]]$$

and let $\tilde{\kappa}_I : T_I \to \widetilde{\Lambda}_I^\times$ be the universal character.

We define a sheaf $\Omega_0^{\tilde{\kappa}_I}$ over $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ by the formula:

$$\Omega_0^{\tilde{\kappa}_I} = ((\pi_* \mathcal{O}_{\mathfrak{IG}^I}) \hat{\otimes}_{\mathbf{Z}_p} \widetilde{\Lambda}_I)^{T_I}$$

where $\pi : \mathfrak{IG}^I \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ is the affine projection, and the group T_I acts diagonally (via its natural action on $\pi_*\mathcal{O}_{\mathfrak{IG}^I}$, and via $\tilde{\kappa}_I$ on $\widetilde{\Lambda}_I$).

We set $\Omega^{\tilde{\kappa}_I} = \Omega_0^{\tilde{\kappa}_I} \bigotimes_{v|p} \det{}^2 \omega_{\mathcal{G}_v}$. The explanation for the twist by this invertible sheaf is given below in §4.3.6 (see in particular Lemma 4.3.8). This is an invertible sheaf of $\mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)} \hat{\otimes}_{\mathbf{Z}_p} \widetilde{\Lambda}_I$-modules.

4.3.6. *Comparison with classical sheaves.* Over $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$ we have for all $v|p$ a surjective map $\omega_{\mathcal{G}_v} \to \omega_{H_{\infty,v}}$ arising from the differential of the inclusion $H_{\infty,v} \hookrightarrow \mathcal{G}_v$.

Let $\kappa = ((k_v, 2)_{v\in I}, (k_v, l_v)_{v\in I^c}) \in (\mathbf{Z}^2)^{S_p}$, $k_v \geq 2$ if $v \in I$, $k_v \geq l_v$ if $v \in I^c$, be an algebraic weight. By construction there is a surjective map

(4.3.7)
$$\omega^\kappa|_{\mathfrak{X}^I_{K_p,\mathrm{Kli}}(p^\infty)} \to (\otimes_{v\in I}\omega_{H_{\infty,v}}^{k_v-2}) \bigotimes (\otimes_{v\in I^c}\omega_{\mathcal{G}_v/H_{\infty,v}}^{l_v-2} \otimes \omega_{H_{\infty,v}}^{k_v-2}) \bigotimes (\otimes_{v|p} \det \omega^2_{\mathcal{G}_v})$$

This map should be interpreted as the projection to the highest weight vector. Moreover over $\mathfrak{IG}^I$ the Hodge–Tate map (see [Mes72, p. 117], as well as Section 6.1.4 below) provides maps:

$$\mathbf{Z}_p \overset{\psi_v}{\to} T_p(H^D_{\infty,v}) \overset{\mathrm{HT}}{\to} \omega_{H_{\infty,v}}$$

for all $v \in S_p$, and

$$\mathbf{Z}_p \overset{\phi_v}{\to} T_p((\mathcal{G}_v^m/H_{\infty,v})^D) \overset{\mathrm{HT}}{\to} \omega_{\mathcal{G}_v/H_{\infty,v}}$$

for all $v \in I^c$.

These maps induce isomorphisms after tensoring with $\mathcal{O}_{\mathfrak{IG}^I}$ on the left. Therefore the Hodge–Tate map provides a $\mathbf{Z}_p^\times$-reduction of the GL_1-torsors $\omega_{H_{\infty,v}}$ and $\omega_{\mathcal{G}_v/H_{\infty,v}}$.

Let $\kappa = ((k_v, 2)_{v\in I}, (k_v, l_v)_{v\in I^c})$ be a classical algebraic weight. Then we can naturally identify κ with a p-adic weight (that is, an element of $\mathrm{Hom}(\widetilde{\Lambda}_I, \mathbf{Z}_p)$) via the character:

$$((x_v)_{v\in I}, (x_v, y_v)_{v\in I^c}) \in T_I \mapsto \prod_{v\in I} x_v^{k_v-2} \prod_{v\in I^c} x_v^{k_v-2} y_v^{l_v-2}$$

Let us define $\Omega^\kappa = \Omega^{\tilde{\kappa}_I} \otimes_{\widetilde{\Lambda}_I,\kappa} \mathbf{Z}_p$.

Lemma 4.3.8. *For all $\kappa = ((k_v, 2)_{v\in I}, (k_v, l_v)_{v\in I^c}) \in (\mathbf{Z}^2)^{S_p}$ with $k_v \geq 2$ if $v \in I$, $k_v \geq l_v$ if $v \in I^c$, there is a canonical isomorphism*

$$\mathrm{HT}^* : (\otimes_{v\in I}\omega_{H_{\infty,v}}^{k_v-2}) \bigotimes (\otimes_{v\in I^c}\omega_{\mathcal{G}_v/H_{\infty,v}}^{l_v-2} \otimes \omega_{H_{\infty,v}}^{k_v-2}) \bigotimes (\otimes_{v|p} \det \omega^2_{\mathcal{G}_v}) \simeq \Omega^\kappa.$$

Proof. By definition, it suffices to construct an isomorphism

$$\mathrm{HT}^* : (\otimes_{v\in I}\omega_{H_{\infty,v}}^{k_v-2}) \bigotimes (\otimes_{v\in I^c}\omega_{\mathcal{G}_v/H_{\infty,v}}^{l_v-2} \otimes \omega_{H_{\infty,v}}^{k_v-2}) \simeq \Omega_0^\kappa$$

where $\Omega_0^\kappa = \Omega_0^{\tilde{\kappa}_I} \otimes_{\widetilde{\Lambda}_I,\kappa} \mathbf{Z}_p$. Sections of the sheaf Ω_0^κ are rules f associating to $(x, (\phi_v)_{v\in I^c}, (\psi_v)_{v\in S_p}) \in \mathfrak{IG}^I(R)$, an element

$$f(x, (\phi_v), (\psi_v)) \in R$$

such that

$$\begin{aligned} f(x, (\phi_v \circ \lambda_v^{-1}), (\psi_v \circ \beta_v^{-1})) &= \kappa((\lambda_v, \beta_v)) f(x, (\phi_v), (\psi_v)) \\ &= \prod_{v\in S_p} \lambda_v^{k_v-2} \prod_{v\in I^c} \beta_v^{l_v-2} f(x, (\phi_v), (\psi_v)) \end{aligned}$$

for $((\lambda_v)_{v\in S_p}, (\beta_v)_{v\in I^c}) \in T_I$.

Sections of the sheaf

$$(\otimes_{v\in I}\omega_{H_{\infty,v}}^{k_v-2})\bigotimes(\otimes_{v\in I^c}\omega_{\mathcal{G}_v/H_{\infty,v}}^{l_v-2}\otimes\omega_{H_{\infty,v}}^{k_v-2})$$

are rules g associating to triples

$$(x, (a_v)_{v|p}, (b_v)_{v\in I^c})$$

for R a p-adically complete $\mathbf{Z}_p$-algebra, $x \in \mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)(R)$, $a_v : R \simeq x^*\omega_{H_{v,\infty}}$, $b_v : R \simeq x^*\ \omega_{\mathcal{G}_v/H_{v,\infty}}$ an element

$$f(x, a_v, b_v) \in R$$

such that

$$f(x, a_v\circ\lambda_v^{-1}, b_v\circ\delta_v^{-1}) = \prod_{v|p}\lambda_v^{k_v-2}\prod_{v\in I^c}\delta_v^{l_w-2}f(x, a_v, b_v)$$

for all $(\lambda_v) \in (R^\times)^{S_p}, (\delta_v) \in (R^\times)^{I^c}$.

To a rule g as above, we associate a rule

$$\mathrm{HT}^*(g)(x, (\phi_v)_{v\in I^c}, (\psi_v)_{v\in S_p}) = g(x, (\mathrm{HT}(\phi_v(1)))_{v\in I^c}, (\mathrm{HT}(\psi_v(1)))_{v\in S_p}).$$

It is easy to check that the map HT^* is an isomorphism. □

We can now summarize the interpolation property of the sheaf $\Omega^{\tilde{\kappa}_I}$.

Corollary 4.3.9. *For all $\kappa = ((k_v, 2)_{v\in I}, (k_v, l_v)_{v\in I^c}) \in (\mathbf{Z}^2)^{S_p}$ with $k_v \geq 2$ if $v \in I$, $k_v \geq l_v$ if $v \in I^c$, there is a canonical surjective map:*

$$\omega^\kappa|_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)} \to \Omega^\kappa = \Omega^{\tilde{\kappa}_I}\otimes_{\tilde{\Lambda}_I,\kappa}\mathbf{Z}_p.$$

Proof. In view of Lemma 4.3.8, this is just the map (4.3.7). □

4.4. **Sheaves of p-adic modular forms for G_1.** In this section we explain how we can descend our construction to the Shimura variety for G_1. This section is the analogue for p-adic sheaves of §3.7.2.

4.4.1. *Weight space for G_1.* We now assume that $p \neq 2$. We let T_I^0 be the pro-p sub-group of T_I, so that $T_I = T_I^f \times T_I^0$ is the product of a finite group T_I^f and T_I^0. We let $\Lambda_I = \mathbf{Z}_p[[T_I^0]]$. There is a canonical projection $T_I \to T_I^0$ and a canonical character $\kappa_1 : T_I \to \Lambda_I^\times$ which identifies Λ_I with the deformation space of the trivial character of T_I. This canonical projection makes Λ_I a quotient of $\tilde{\Lambda}_I$. We let $\kappa_I = \kappa_1\otimes((2, 2)_{v|p})$. The pair (κ_I, Λ_I) is the universal deformation space of the character $((2, 2)_{v|p})$ mod p. We let $\Omega^{\kappa_I} = \Omega^{\tilde{\kappa}_I}\otimes_{\tilde{\Lambda}_I}\Lambda_I$.

If $\kappa = ((k_v, 2)_{v\in I}, (k_v, l_v)_{v\in I^c})$ is a classical algebraic weight such that $k_v \equiv l_v \equiv 2 \pmod{p-1}$, then κ defines a $\mathbf{Z}_p$-point of Spf Λ_I in the following way: we associate to κ the character

$$((x_v)_{v\in I}, (x_v, y_v)_{v\in I^c}) \in T_I \mapsto \prod_{v\in I}x_v^{k_v-2}\prod_{v\in I^c}x_v^{k_v-2}y_v^{l_v-2}$$

which factors through a character of T_I^0 and therefore defines a morphism $f_\kappa : \Lambda_I \to \mathbf{Z}_p$. The specialization of κ_I along the map f_κ recovers the character κ.

4.4.2. *Descent.* The group $(\mathcal{O}_F)^{\times,+}_{(p)}$ can be embedded "diagonally" in T_I by sending $x \in (\mathcal{O}_F)^{\times,+}_{(p)}$ to $((x_v)_v \in I, (x_v, x_v)_{v \in I^c})$ where for all places $v|p$ we denote by $x_v \in \mathcal{O}^\times_{F_v} = \mathbf{Z}_p^\times$ the image of x in F_v. For an element $x \in T_I$, we denote by x_0 the projection of x to T_I^0. For an element $x \in T_I^0$, we denote by $\tilde{x}$ the corresponding group element in Λ_I.

Since T_I^0 is a pro-p group, and $p > 2$, the map $x \mapsto x^2$ is bijective on T_I^0. Accordingly if $x \in T_I^0$, then we define $\sqrt{x} \in T_I^0$ by the equation $(\sqrt{x})^2 = x$. We then define a character $d : (\mathcal{O}_F)^{\times,+}_{(p)} \to (\Lambda_I)^\times$ (where "d" stands for "descent") by the formula:

$$d(x) = \sqrt{x_0}.$$

The group $(\mathcal{O}_F)^{\times,+}_{(p)}$ acts on $\mathfrak{X}^I_K$; in the notation of §3.3, the element $x \in (\mathcal{O}_F)^{\times,+}_{(p)}$ sends $(A, \iota, \lambda, \eta = (\eta_1, \eta_2), \eta_p)$ to $(A, \iota, x\lambda, (\eta_1, x\eta_2), \eta_p)$. We can lift this action to Ω^{κ_I} by setting

$$x : x^*\Omega^\kappa_I \to \Omega^\kappa_I$$

to be the composition of the tautological isomorphism (the construction of Ω^{κ_I} doesn't depend on the polarization) and multiplication by $d(x)$. The reader can easily check that this defines an action and is compatible with the construction of §3.7.2; as always, we are making the choice $w = 2$.

For all $n \in \mathbf{Z}_{\geq 0} \cup \{\infty\}$, we can form the quotient of $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^n)$ by the action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ (which factors through a finite group acting freely) and we denote by $\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^n)$ the corresponding quotient. The maps $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^n) \to \mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^n)$ are étale. We can also descend the sheaf Ω^{κ_I} to a sheaf Ω^{κ_I} over $\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty)$ using the descent datum provided by d.

We let $M_I^{p\text{-}ad,\kappa_I} = \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I}(-D))$ be the cohomology of the p-adic cuspidal modular forms of weight κ_I.

Proposition 4.4.3. *The canonical map $M_I^{p\text{-}ad,\kappa_I} \to \mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I}(-D))$ is split in the derived category of $\mathbf{Z}_p$-modules.*

Proof. See the proof of Proposition 3.8.3. □

4.5. **Hecke operators at p on the cohomology of p-adic modular forms.** In this section we define Hecke operators at p acting on the cohomology of p-adic modular forms. Recall that we have fixed a partition $S_p = \{v|p\} = I \coprod I^c$.

4.5.1. *Hecke operators of Siegel type.* Let $w \in I^c$ be a place above p. Let $K = K^pK_p$ be a reasonable compact open subgroup with $K_p = G_1(\mathbf{Z}_p)$. In §3.9.17 we defined a Hecke operator attached to the correspondence (for suitable choices of polyhedral cone decompositions omitted from the notation):

$$\begin{tikzcd} & X_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_K & & X_K \end{tikzcd}$$

where $K' = K^pK'_p$ and $K'_p = \prod_{v|p, v\neq w} \mathrm{GSp}_4(\mathcal{O}_{F_v}) \times \mathrm{Si}(w)$. The map p_2 depends on the choice of an element $x_w \in F^{\times,+}$. We are now going to pull back this correspondence to a deep Klingen level structure and isolate the "essential part".

As in Remark 3.9.23, the resulting Hecke operators are easily seen to be independent of the choice of polyhedral cone decomposition.

Taking formal p-adic completions, we obtain a correspondence:

$$\begin{tikzcd} & \mathfrak{X}_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ \mathfrak{X}_K & & \mathfrak{X}_K \end{tikzcd} \tag{4.5.2}$$

We consider the fibre product $\mathfrak{X}_{K'} \times_{p_1, \mathfrak{X}_K} \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. Recall that by our assumption that $w \in I^c$, $\mathcal{G}_w$ is an ordinary Barsotti–Tate group. We denote by $\mathfrak{C}_{w,1}(p^m)$ the open and closed formal subscheme of this fibre product where the kernel of the canonical isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ has trivial multiplicative part (it is open and closed by the rigidity of multiplicative groups).

There is an obvious map $u_1 : \mathfrak{C}_{w,1}(p^m) \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$, given by projection onto the second factor of the fibre product. We claim that the projection $\mathfrak{C}_{w,1}(p^m) \to \mathfrak{X}_K$ induced by p_2 can be lifted to a map $u_2 : \mathfrak{C}_{w,1}(p^m) \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. Indeed, since H_m is multiplicative, the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ induces an isomorphism from $p_1^*H_m$ to its image in $p_2^*\mathcal{G}$. We call this image $p_2^*H_m$. We therefore have a correspondence

$$\begin{tikzcd} & \mathfrak{C}_{w,1}(p^m) \arrow[dl, "u_2"'] \arrow[dr, "u_1"] & \\ \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) & & \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) \end{tikzcd}$$

We now associate to this correspondence a Hecke operator $U_{w,1}$.

Remark 4.5.3. The Hecke operator $U_{w,1}$ is the standard "U_p" operator (at the place w) that is considered in the usual theory of p-adic modular forms.

Lemma 4.5.4. *There is a normalized trace map* "$\frac{1}{p^3}\mathrm{Tr}_{u_1}$" : $\mathrm{R}(u_1)_*\mathcal{O}_{\mathfrak{C}_{w,1}(p^m)} \to \mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)}$.

Proof. The formal schemes $\mathfrak{C}_{w,1}(p^m)$ and $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ are smooth over $\mathbf{Z}_p$. Consider the map induced by u_1 on top-differentials:

$$\mathrm{d}u_1 : \det \Omega^1_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)/\mathbf{Z}_p} \to \det \Omega^1_{\mathfrak{C}_{w,1}(p^m)/\mathbf{Z}_p}.$$

This map is divisible by p^3 by the same arguments as in the proof of Lemma 3.9.18. Namely, the map u_1 is totally inseparable and hence a homeomorphism. For any closed point $x \in \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ in the interior, one sees by Serre–Tate theory that the map of completed local rings $\widehat{\mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m),x}} \to \widehat{\mathcal{O}_{\mathfrak{C}_{w,1}(p^m),x}}$ is given by $\otimes_{v|p} W(k(x))[[T_{1,v}, T_{2,v}, T_{3,v}]] \to \otimes_{v|p} W(k(x))[[T_{1,v}, T_{2,v}, T_{3,v}]]$ where $T_{i,v} \mapsto T_{i,v}$ if $v \neq w$, and $T_{i,w} \mapsto (1 + T_{i,w})^p - 1$.

By reduction modulo p^n of u_1, we get a proper map $u_1 : C_{w,1}(p^m)_n \to X^I_{K,\mathrm{Kli}}(p^m)_n$ of smooth schemes over $\mathrm{Spec}\, \mathbf{Z}/p^n\mathbf{Z}$. The above map $\frac{1}{p^3}\mathrm{d}u_1$ induces a map $\mathcal{O}_{C_{w,1}(p^m)_n} \to u_1^!\mathcal{O}_{X^I_{K,\mathrm{Kli}}(p^m)_n}$ or by adjunction a map "$\frac{1}{p^3}\mathrm{Tr}_{u_1}$" : $\mathrm{R}(u_1)_*(\mathcal{O}_{\mathfrak{C}_{w,1}(p^m)}/p^n) \to (\mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)})/p^n$ (see §3.8.11). Passing to the limit over n yields the map of the lemma. □

Remark 4.5.5. We sketch another argument for the proof of the lemma. Write $\mathfrak{C}_{w,1}(p^m)_{\mathfrak{Y}}$ for the restriction of $\mathfrak{C}_{w,1}(p^m)$ to $\mathfrak{Y}^I_{K,\mathrm{Kli}}(p^m)$. It is easy to show that

the map $u_1 : \mathfrak{C}_{w,1}(p^m)_{\mathfrak{Y}} \to \mathfrak{Y}^I_{K,\mathrm{Kli}}(p^m)$ is finite flat of degree p^3. Therefore, the restriction of the map of the lemma to $\mathfrak{Y}^I_{K,\mathrm{Kli}}(p^m)$ is the usual trace map, normalized by a factor p^{-3}. Let Σ be the K-admissible polyhedral cone decomposition such that $\mathfrak{X}_K = \mathfrak{X}_{K,\Sigma}$ and $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) = \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)_\Sigma$. We can use the same Σ to get the (non smooth) toroidal compactification of $\mathfrak{X}_{K',\Sigma}$ and then of $\mathfrak{C}_{w,1}(p^m)_\Sigma$. Now we observe that the map $\mathfrak{C}_{w,1}(p^m)_\Sigma \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)_\Sigma$ is finite flat and therefore has a trace map. It remains to recall that for any refinement Σ' of Σ, the map $\pi : \mathfrak{C}_{w,1}(p^m)_{\Sigma'} \to \mathfrak{C}_{w,1}(p^m)_\Sigma$ induces a quasi-isomorphism: $\mathrm{R}\pi_*\mathcal{O}_{\mathfrak{C}_{w,1}(p^m)_{\Sigma'}} = \mathcal{O}_{\mathfrak{C}_{w,1}(p^m)_\Sigma}$.

To define the Hecke operator $U_{w,1}$ on $\mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I})$, $\mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m), \omega^\kappa)$, and so on, we argue as follows. By the usual formalism, if $\mathcal{F}$ is one of Ω^{κ_I} or ω^κ, it is enough to define morphisms $u_2^*\mathcal{F} \to u_1^*\mathcal{F}$; we can then compose with the trace map of Lemma 4.5.4.

To this end, note that over $\mathfrak{C}_{w,1}(p^m)$ we have the canonical étale isogeny $u_1^*\mathcal{G} \to u_2^*\mathcal{G}$, which determines an isomorphism on differentials. We thus have a canonically determined isomorphism $u_2^*\omega^\kappa \to u_1^*\omega^\kappa$ (with no need to normalize). Similarly, since the canonical isogeny induces an isomorphism $u_1^*H_m \to u_2^*H_m$, we have a canonical isomorphism $u_2^*\Omega^{\kappa_I} \to u_1^*\Omega^{\kappa_I}$ (again with no need to normalize).

4.5.6. *The operator* $U_{\mathrm{Kli}(w),1}$. We now introduce another Hecke operator of Siegel type for $w \in I$, which we denote $U_{\mathrm{Kli}(w),1}$. This operator will not be used until §5 and §7. We decided to introduce it here because its definition is similar to the other operators of Siegel type introduced in §4.5.1, and because it is convenient to discuss the commutativity of all of our Hecke operators at p in one go (see Lemma 4.5.15 below). We defer the details of the normalization of this Hecke operator to §5.3, where we will consider the operator $U_{\mathrm{Kli}(w),1}$ in a more general context.

We again consider the correspondence (4.5.2), and the product $\mathfrak{X}_{K'} \times_{p_1,\mathfrak{X}_K} \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. We denote by $\mathfrak{C}_{\mathrm{Kli}(w),1}(p^m)$ the open and closed formal subscheme of this fibre product where the kernel of the canonical isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ has trivial intersection with the group $p_1^*H_m$. Exactly as above, we obtain a correspondence

$$\begin{tikzcd} & \mathfrak{C}_{\mathrm{Kli}(w),1}(p^m) \arrow[dl, "v_2"'] \arrow[dr, "v_1"] & \\ \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) & & \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) \end{tikzcd}$$

We show in Lemma 5.3.2 below that there is a Hecke operator

$$U_{\mathrm{Kli}(w),1} : \mathrm{R}(v_1)_* v_2^*\omega^2 \to \omega^2, \tag{4.5.7}$$

defined using a trace map normalized by a factor of $1/p^3$. On the other hand, we have natural isomorphisms $v_2^*\Omega_0^{\tilde{\kappa}_I} \to v_1^*\Omega_0^{\tilde{\kappa}_I}$, and tensoring this map with (4.5.7) produces the desired cohomological correspondence (and associated Hecke operator):

$$U_{\mathrm{Kli}(w),1} : \mathrm{R}(v_1)_* v_2^*\Omega^{\kappa_I} \to \Omega^{\kappa_I}.$$

4.5.8. *Hecke operators of Klingen type.* Let $w|p$ be a place. Let $K = K^pK_p$ be a reasonable compact open subgroup with $K_p = G_1(\mathbf{Z}_p)$. In §3.9.20 we have defined

a Hecke operator attached to the correspondence (again, for suitable choices of polyhedral cone decompositions omitted from the notation):

$$\begin{tikzcd} & X_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ X_{K''} & & X_K \end{tikzcd}$$

where $K' = K^p K'_p$ with $K'_p = \prod_{v|p, v\neq w} \mathrm{GSp}_4(\mathcal{O}_{F_v}) \times \mathrm{Kli}(w)$, and $K'' = K^p K''_p$ with $K''_p = \prod_{v|p, v\neq w} \mathrm{GSp}_4(\mathcal{O}_{F_v}) \times \mathrm{Par}(w)$. The map p_2 depends on the choice of an element $x_w \in F^{\times,+}$, and over $X_{K'}$ we have natural isogenies

$$p_1^*\mathcal{G} \to p_2^*\mathcal{G} \to p_1^*\mathcal{G}$$

whose composite is multiplication by x_w. As in §4.5.1, we are going to pull back this correspondence to a deep Klingen level structure and isolate the "essential part" of the correspondence.

Taking formal p-adic completions, we obtain:

$$\begin{tikzcd} & \mathfrak{X}_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ \mathfrak{X}_{K''} & & \mathfrak{X}_K \end{tikzcd}$$

We consider the fibre product $\mathfrak{X}_{K'} \times_{p_1, \mathfrak{X}_K} \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. We denote by $\mathfrak{C}_{w,2,1}(p^m)$ the formal subscheme where the kernel of the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ has trivial intersection with the group $p_1^* H_m$.

Lemma 4.5.9. *The formal subscheme $\mathfrak{C}_{w,2,1}(p^m)$ is open and closed in $\mathfrak{X}_{K'} \times_{p_1, \mathfrak{X}_K} \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$.*

Proof. Let $L = p_1^* H_m \cap \mathrm{Ker}(p_1^*\mathcal{G} \to p_2^*\mathcal{G})$. Since L is a closed subscheme of $p_1^* H_m$, it is finite over $\mathfrak{X}_{K'} \times_{p_1, \mathfrak{X}_K} \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ and the condition that $L = \{0\}$ is therefore open. It is also closed because if at some point x there is a non-trivial map $p_1^* H_m|_x \to \mathrm{Ker}(p_1^*\mathcal{G} \to p_2^*\mathcal{G})|_x$, this map will extend on the completed local ring at x by the rigidity of multiplicative groups. □

There is an obvious map $r_1 : \mathfrak{C}_{w,2,1}(p^m) \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ induced by the projection p_1. We claim that the second projection p_2, which induces a map $\mathfrak{C}_{w,2,1}(p^m) \to \mathfrak{X}_{K''}$, can be lifted to a map $r_2 : \mathfrak{C}_{w,2,1}(p^m) \to \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m)$. Indeed, over $\mathfrak{C}_{w,2,1}(p^m)$ the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ induces an isomorphism from $p_1^* H_m$ to its image in $p_2^*\mathcal{G}$ (which we call $p_2^* H_m$). We therefore have a correspondence

$$\begin{tikzcd} & \mathfrak{C}_{w,2,1}(p^m) \arrow[dl, "r_2"'] \arrow[dr, "r_1"] & \\ \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m) & & \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) \end{tikzcd}$$

We now associate to this correspondence a Hecke operator

$$U'_w \in \mathrm{Hom}(\mathrm{R}\Gamma(\mathfrak{X}^I_{K'',\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I}), \mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I})).$$

To do so, we have the following lemma.

Lemma 4.5.10. *There is a normalized trace map "$\frac{1}{p^2}\mathrm{Tr}_{r_1}$": $(r_1)_*\mathcal{O}_{\mathfrak{C}_{w,2,1}(p^m)} \to \mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)}$*

Proof. The formal schemes $\mathfrak{C}_{w,2,1}(p^m)$ and $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ are smooth over $\mathbf{Z}_p$. Consider the induced map on top differentials:

$$\mathrm{d}r_1 : \det \Omega^1_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)/\mathbf{Z}_p} \to \det \Omega^1_{\mathfrak{C}_{w,2,1}(p^m)/\mathbf{Z}_p}$$

This map is divisible by p^2 for the same reason as in the proof of Lemma 3.9.22. Namely, let us fix a closed point $x \in \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$ which is in the interior and ordinary at w. The fibre of r_1 at x parametrizes the subgroup $L = \mathrm{Ker}(p_1^*\mathcal{G} \to p_2^*\mathcal{G})$ of $\mathcal{G}_w[p]$ of étale rank 2, multiplicative rank 1 and trivial intersection with H_m. The total degree of r_1 is p^3. The fibre of r_1 over x has p points (corresponding to the choice of the multiplicative part L^m of L). The inseparability degree is p^2 (corresponding to finding sections of $\mathcal{G}_w[p]/L^m \to \mathcal{G}_w[p]^{et}$). For any $x' \in \mathfrak{C}_{w,2,1}(p^m)$ lying above x, Serre–Tate theory shows that the map on completed local rings $\widehat{\mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m),x}} \to \widehat{\mathcal{O}_{\mathfrak{C}_{w,2,1}(p^m),x'}}$ is isomorphic to $\otimes_{v|p} W(k(x))[[T_{1,v}, T_{2,v}, T_{3,v}]] \to \otimes_{v|p} W(k(x'))[[T_{1,v}, T_{2,v}, T_{3,v}]]$ where $T_{i,v} \mapsto T_{i,v}$ if $v \neq w$ or $i = 1$, and $T_{i,w} \mapsto (1 + T_{i,w})^p - 1$ for $i = 2, 3$.

By reduction modulo p^n of r_1, we get a proper map $r_1 : C_{w,2,1}(p^m)_n \to X^I_{K,\mathrm{Kli}}(p^m)_n$ of smooth schemes over $\mathrm{Spec}\, \mathbf{Z}/p^n\mathbf{Z}$. The above map $\frac{1}{p^2}\mathrm{d}r_1$ induces a map $\mathcal{O}_{C_{w,2,1}(p^m)_n} \to r_1^!\mathcal{O}_{X^I_{K,\mathrm{Kli}}(p^m)_n}$ or by adjunction a map "$\frac{1}{p^2}\mathrm{Tr}_{r_1}$" : $\mathrm{R}(r_1)_*(\mathcal{O}_{\mathfrak{C}_{w,2,1}(p^m)}/p^n) \to (\mathcal{O}_{\mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)})/p^n$ (see §3.8.11). Passing to the limit over n yields the map of the lemma. □

Remark 4.5.11. We could give an alternative proof of Lemma 4.5.10 as in Remark 4.5.5.

As usual over $\mathfrak{C}_{w,2,1}(p^m)$ we have the canonical isogeny $r_1^*\mathcal{G} \to r_2^*\mathcal{G}$, whose differential determines a morphism $r_2^*\omega^\kappa \to r_1^*\omega^\kappa$. We have a commutative diagram

$$\begin{array}{ccccc} r_2^*\omega_{\mathcal{G}_w} & \longrightarrow & r_2^*\omega_{H_{m,w}} & \longrightarrow & 0 \\ \downarrow & & \downarrow & & \\ r_1^*\omega_{\mathcal{G}_w} & \longrightarrow & r_1^*\omega_{H_{m,w}} & \longrightarrow & 0 \end{array}$$

which Zariski locally on affine opens $\mathrm{Spf}\, R$ is isomorphic to

$$\begin{array}{ccccc} & R^2 & \xrightarrow{(0\ 1)} & R & \longrightarrow 0 \\ \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} & \downarrow & & \downarrow{\scriptstyle 1_R} & \\ & R^2 & \xrightarrow{(0\ 1)} & R & \longrightarrow 0 \end{array} \tag{4.5.12}$$

It follows that we can and do normalize the morphism $r_2^*\omega^\kappa \to r_1^*\omega^\kappa$ by dividing by p^{l_w}. When $m = \infty$, the isogeny induces an isomorphism $r_1^*H_{\infty,w} \to r_2^*H_{\infty,w}$, and we therefore obtain an isomorphism $r_2^*\Omega^{\kappa_I} \to r_1^*\Omega^{\kappa_I}$. Combining this with Lemma 4.5.10 gives the desired operator U'_w.

We now exchange the roles of p_1 and p_2, and consider the fibre product $\mathfrak{X}_{K'} \times_{p_2,\mathfrak{X}''_K} \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m)$. We denote by $\mathfrak{C}_{w,2,2}(p^m)$ the open and closed formal subscheme where the kernel of the isogeny $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ has trivial connected component (so that away from the boundary, the kernel of this isogeny is étale). Note that by definition

this kernel is contained in the kernel of the quasi-polarization $p_2^*\mathcal{G} \to p_2^*\mathcal{G}^D$, so the kernel of $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ has multiplicative rank at least 1.

The projection p_2 induces a map $s_2 : \mathfrak{C}_{w,2,2}(p^m) \to \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m)$. We claim that the first projection $p_1 : \mathfrak{C}_{w,2,2}(p^m) \to \mathfrak{X}_K$ can be lifted to a map $s_1 : \mathfrak{C}_{w,1,2}(p^m) \to \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m)$. Indeed, since H_m is connected, we see that over $\mathfrak{C}_{w,2,2}(p^m)$ the isogeny $p_2^*\mathcal{G} \to p_1^*\mathcal{G}$ induces an isomorphism from $p_2^*H_m$ to its image in $p_1^*\mathcal{G}$ (which we call $p_1^*H_m$); and the map $\mathfrak{C}_{w,2,2}(p^m) \to \mathfrak{X}_K$ factors through $\mathfrak{X}^I_K$. Accordingly, we have a correspondence

$$\begin{tikzcd} & \mathfrak{C}_{w,2,2}(p^m) \arrow[dl, "s_1"'] \arrow[dr, "s_2"] & \\ \mathfrak{X}^I_{K,\mathrm{Kli}}(p^m) & & \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m) \end{tikzcd}$$

We can associate to this correspondence a Hecke operator

$$U''_w \in \mathrm{Hom}(\mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I}), \mathrm{R}\Gamma(\mathfrak{X}^I_{K'',\mathrm{Kli}}(p^\infty), \Omega^{\kappa_I})),$$

which again depends on the construction of a trace map:

Lemma 4.5.13. *There is a normalized trace map* $p^{-1}\,\mathrm{Tr}_{s_2} : \mathrm{R}(s_2)_*\mathcal{O}_{\mathfrak{C}_{w,2,2}(p^m)} \to \mathcal{O}_{\mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m)}$.

Proof. This is a calculation in Serre–Tate theory which is similar to the proof of Lemma 4.5.10. Namely, let us fix a closed point $x \in \mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m)$ which is in the interior. The fibre of s_2 at x parametrizes rank p étale subgroups in the kernel of the quasi-polarization $\mathcal{G}_w \to \mathcal{G}_w^D$ (which is a rank p^2 finite flat group scheme, extension of an étale by a multiplicative subgroup). We deduce that the map s_2 is totally inseparable at x of degree p. Serre–Tate theory shows that the map on completed local rings $\widehat{\mathcal{O}_{\mathfrak{X}^I_{K'',\mathrm{Kli}}(p^m),x}} \to \widehat{\mathcal{O}_{\mathfrak{C}_{w,2,2}(p^m),x}}$ is isomorphic to $\otimes_{v|p}W(k(x))[[T_{1,v},T_{2,v},T_{3,v}]] \to \otimes_{v|p}W(k(x))[[T_{1,v},T_{2,v},T_{3,v}]]$ where $T_{i,v} \mapsto T_{i,v}$ if $v \neq w$ or $i = 1,2$, and $T_{3,w} \mapsto (1+T_{3,w})^p - 1$. □

Over $\mathfrak{C}_{w,2,2}(p^m)$ we have the canonical isogeny $s_2^*\mathcal{G} \to s_1^*\mathcal{G}$, which is étale, and therefore determines isomorphisms $s_1^*\omega^\kappa \to s_2^*\omega^\kappa$ and $s_1^*\Omega^{\kappa_I} \to s_2^*\Omega^{\kappa_I}$. Combining with Lemma 4.5.13, we get the desired Hecke operator U''_w.

We set $U_{w,2} := U'_w \circ U''_w$.

4.5.14. *Commutativity of the Hecke operators.* We remind the reader that whenever we write $U_{v,1}$ below, we mean $U_{\mathrm{Iw}(v),1}$.

Lemma 4.5.15. *The operators* $\{U_{\mathrm{Kli}(v),1}, U_{v,2}\}_{v\in I}$ *and* $\{U_{v,1}, U_{v,2}\}_{v\in I^c}$ *commute with each other on* $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}(p^\infty)}, \Omega^{\kappa_I}(-D))$.

Proof. We prove this in the same way as Lemma 4.2.36. We first introduce a similar complex as in §4.2.30 to compute the cohomology. Namely, there is a complex $L^\bullet$ computing $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}(p^\infty)}, \Omega^{\kappa_I}(-D))$, such that $L^k = \varprojlim_t L^k_t$ where L^k_t is

$$\bigoplus_{J\subset I,\#J=k} \varinjlim_{\times\prod_{w\in J}\mathrm{Ha}(\mathcal{G}_w)} H^0\big(\mathfrak{X}^{G_1,I,\geq_{J^c}2}_{K,\mathrm{Kli}(p^\infty)}, \mathbf{Z}/p^t\mathbf{Z}\otimes\Omega^{\kappa_I}(-D)\otimes\bigotimes_{w\in J}(\det\mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n)\big)$$

Note that in that formula we use the fact that $\mathrm{Ha}(\mathcal{G}_w)^n$ has a canonical lift to $\mathbf{Z}/p^t\mathbf{Z}$ for all n's which are multiples of p^{t-1} and these are cofinal among all natural numbers.

Each term

(4.5.16)
$$\varinjlim_{\times \prod_{w\in J} \mathrm{Ha}(\mathcal{G}_w)} H^0\big(\mathfrak{X}^{G_1,I,\geq_{J^c}2}_{K,\mathrm{Kli}(p^\infty)}, \mathbf{Z}/p^t\mathbf{Z}\otimes\Omega^{\kappa_I}(-D)\otimes\bigotimes_{w\in J}(\det\mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n)\big)$$

appearing in this complex is stable under the Hecke action and it is therefore enough to prove the commutativity for each of these terms. Each of the terms (4.5.16) can be embedded into the corresponding direct limit of cohomology groups taken over the interior of the moduli space. Over the interior of the moduli space all our correspondences are finite flat and the commutativity follows from standard properties of the Iwahori and Klingen Hecke algebras. □

Lemma 4.5.17. *If $w \in I$ then we have an equality of Hecke operators*

$$U_{\mathrm{Iw}(w),1}(U_{\mathrm{Kli}(w),1} - U_{\mathrm{Iw}(w),1}) = U_{w,2}$$

on $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I,=_w 2}_{K,\mathrm{Kli}(p^\infty)}, \Omega^{\kappa_I}(-D))$.

Proof. Using a Cousin complex computing the cohomology as in the proof of Lemma 4.5.15, we reduce to proving that the underlying (cohomological) correspondences agree away from the boundary. By definition, the correspondence associated to $U_{\mathrm{Iw}(w),1}$ parameterizes triples $(\mathcal{G}, H_m, L)$ where $L \subset \mathcal{G}_w[p]$ is étale, totally isotropic of degree p^2, and has $L \cap H_m = \{0\}$. Similarly, the correspondence associated to Z_w parameterizes triples $(\mathcal{G}, H_m, M)$ where $M \subset \mathcal{G}_w[p]$ has multiplicative rank 1, is totally isotropic of degree p^2, and has $M \cap H_m = \{0\}$. Finally, as in [Pil20, Prop. 10.2.1], the correspondence associated to $U_{w,2}$ parameterizes triples $(\mathcal{G}, H_m, N)$ where $N \subset \mathcal{G}_w[p^2]$ is totally isotropic of degree p^4, $N[p]$ has degree p^3, and $N \cap H_m = \{0\}$.

Comparing these definitions, we see that on the level of underlying correspondences, we have $pU_{w,2} = U_{\mathrm{Iw}(w),1}Z_w$. Since the normalization factors involved in the Hecke operators $U_{w,2}, U_{\mathrm{Iw}(w),1}, Z_w$ are respectively p^{-5}, p^{-3}, p^{-3}, and since $5+1 = 3+3$, the result follows. □

4.6. **Perfect complexes of p-adic modular forms.** Let $K = K^pK_p$ be a reasonable compact open subgroup with $K_p = G_1(\mathbf{Z}_p)$. Set

$$U^I = \prod_{v\in I} U_{v,2} \prod_{v\in I^c} U_{v,1}U_{v,2}.$$

This is an endomorphism of $M_I^{p\text{-}ad,\kappa_I} = \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}(p^\infty)}, \Omega^{\kappa_I}(-D))$, an object of the bounded derived category of Λ_I-modules. In this section we prove the following theorem.

Theorem 4.6.1.

(1) The operator U^I is locally finite on $M_I^{p\text{-}ad,\kappa_I}$.

(2) Let $e(U^I)$ be the ordinary projector attached to U^I and let $M_I := e(U^I)M_I^{p\text{-}ad,\kappa_I}$ be the associated direct summand. Then the complex M_I is a perfect complex of Λ_I-modules concentrated in the interval $[0, \#I]$.

(3) For all classical algebraic weights $\kappa = ((k_v, l_v)_{v|p})$ with $l_v = 2$ when $v \in I$ and $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$, there is a canonical quasi-isomorphism:

$$e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K_p(I)K^p}, \omega^\kappa(-D)) \to M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{Z}_p.$$

(4) *There is a universal constant C depending only on p but not on the tame level K^p such that for all classical algebraic weights $\kappa = ((k_v, l_v)_{v|p})$ with $l_v = 2$ when $v \in I$, $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$, $k_v - l_v \geq C$ when $v|p$, and $l_v \geq C$ when $v \in I^c$, the map:*

$$e(\prod_{v|p} T_v \prod_{v \in I^c} T_{v,1}) H^i(X_K^{G_1}, \omega^\kappa(-D)) \to H^i(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{Z}_p)$$

is an isomorphism for $i = 0$ and injective for $i = 1$.

We will deduce the theorem from a number of intermediate results. In particular, we need to analyze the Hecke operators $U_{w,1}$ and $U_{w,2}$ and relate them to $T_{w,1}$ and T_w in order to be able to use the results of §4.2.

4.6.2. *Reduction of the correspondence modulo p.* Let $w \in I^c$. We begin by considering the special fibre of the correspondence over $\mathfrak{X}^I_{K,\mathrm{Kli}(p)}$ underlying the operator $U_{w,1}$; we write $C_{w,1}(p)_1$ for this special fibre. By reduction modulo p, it follows from Lemma 3.8.10 that for each classical algebraic weight κ we obtain a cohomological correspondence which we continue to denote by $U_{w,1} : u_2^*(\omega^\kappa|_{X^I_{K,\mathrm{Kli}}(p)_1}) \to u_1^!(\omega^\kappa|_{X^I_{K,\mathrm{Kli}}(p)_1})$.

Lemma 4.6.3. *For any place $w \in I^c$, we have a commutative diagram*

$$\begin{tikzcd}
u_2^*\omega^\kappa \arrow[r, "U_{w,1}"] \arrow[d, "u_2^*\mathrm{Ha}(\mathcal{G}_w)"] & u_1^!\omega^\kappa \arrow[d, "u_1^*\mathrm{Ha}(\mathcal{G}_w)"] \\
u_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "U_{w,1}"] & u_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})
\end{tikzcd}$$

Proof. Since the kernel of $u_1^*\mathcal{G} \to u_2^*\mathcal{G}$ is étale, and the formation of $\mathrm{Ha}(\mathcal{G}_w)$ commutes with étale isogenies, this is immediate. □

We now consider the operator $U_{w,2}$ on $X^I_{K,\mathrm{Kli}}(p)_1$, where w is any place lying over p. Taking the special fibres of the correspondences of §4.5.8 with $m = 1$, we have a correspondence

$$\begin{tikzcd}
& C_{w,2,1}(p)_1 \arrow[dl, "r_2"'] \arrow[dr, "r_1"] & \\
X^I_{K'',\mathrm{Kli}}(p)_1 & & X^I_{K,\mathrm{Kli}}(p)_1
\end{tikzcd}$$

and by Lemma 3.8.10, a cohomological correspondence $r_2^*\omega^\kappa|_{X^I_{K'',\mathrm{Kli}}(p)_1} \to r_1^!\omega^\kappa|_{X^I_{K,\mathrm{Kli}}(p)_1}$; and a correspondence

$$\begin{tikzcd}
& C_{w,2,2}(p)_1 \arrow[dl, "s_1"'] \arrow[dr, "s_2"] & \\
X^I_{K,\mathrm{Kli}}(p)_1 & & X^I_{K'',\mathrm{Kli}}(p)_1
\end{tikzcd}$$

and again by Lemma 3.8.10, a cohomological correspondence $s_2^*\omega^\kappa|_{X^I_{K,\mathrm{Kli}}(p)_1} \to s_1^!\omega^\kappa|_{X^I_{K'',\mathrm{Kli}}(p)_1}$.

We can associate to these cohomological correspondences Hecke operators which we denote as before as

$$U'_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X^I_{K'',\mathrm{Kli}}(p)_1, \omega^\kappa), \mathrm{R}\Gamma(X^I_{K,\mathrm{Kli}}(p)_1, \omega^\kappa)),$$

$$U''_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X^I_{K,\mathrm{Kli}}(p)_1, \omega^\kappa), \mathrm{R}\Gamma(X^I_{K'',\mathrm{Kli}}(p)_1, \omega^\kappa)).$$

We continue to write $U_{w,2} = U'_w \circ U''_w$.

Lemma 4.6.4. *For any $w|p$, we have commutative diagrams*

$$\begin{tikzcd} r_2^*\omega^\kappa \arrow[r, "U'_w"] \arrow[d, "r_2^*\mathrm{Ha}(\mathcal{G}_w)"] & r_1^!\omega^\kappa \arrow[d, "r_1^*\mathrm{Ha}(\mathcal{G}_w)"] \\ r_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "U'_w"] & r_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \end{tikzcd}$$

$$\begin{tikzcd} s_1^*\omega^\kappa \arrow[r, "U''_w"] \arrow[d, "s_1^*\mathrm{Ha}(\mathcal{G}_w)"] & s_2^!\omega^\kappa \arrow[d, "s_2^*\mathrm{Ha}(\mathcal{G}_w)"] \\ s_1^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \arrow[r, "U''_w"] & s_2^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \end{tikzcd}$$

Proof. See [Pil20, Lem. 10.5.2.1]. □

4.6.5. *Reduction of the correspondences to the non-ordinary locus.*

Lemma 4.6.6. *For any $w|p$, the Hasse invariant $\mathrm{Ha}(\mathcal{G}_w)$ is not a zero divisor on each of $C_{w,2,1}(p)_1$ and $C_{w,2,2}(p)_1$.*

Proof. See [Pil20, Lem. 10.5.2.2]. □

We now assume $w \in I$ (otherwise the schemes we consider would be empty) and consider the rank one locus at w, $X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1$, which by definition is the vanishing locus of $\mathrm{Ha}(\mathcal{G}_w)$ in $X^I_{K,\mathrm{Kli}}(p)_1$. Taking the zero locus of $\mathrm{Ha}(\mathcal{G}_w)$ at all entries of the correspondences $C_{w,1}(p)_1$, $C_{w,2,2}(p)_1$ and $C_{w,2,1}(p)_1$ (and taking into account Lemmas 4.6.3 and 4.6.4), we obtain correspondences

$$\begin{tikzcd} & C^{=_w 1}_{w,1}(p)_1 \arrow[dl, "u_2"'] \arrow[dr, "u_1"] & \\ X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1 & & X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1 \end{tikzcd}$$

$$\begin{tikzcd} & C^{=_w 1}_{w,2,1}(p)_1 \arrow[dl, "r_2"'] \arrow[dr, "r_1"] & \\ X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1 & & X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1 \end{tikzcd}$$

$$\begin{tikzcd} & C^{=_w 1}_{w,2,2}(p)_1 \arrow[dl, "s_1"'] \arrow[dr, "s_2"] & \\ X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1 & & X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1 \end{tikzcd}$$

By Lemmas 3.8.10 and 4.6.6, we also obtain cohomological correspondences

$$u_2^*\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli},1}(p)} \to u_1^!\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1}, \quad r_2^*\omega^\kappa|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1} \to r_1^!\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1},$$

and

$$s_2^*\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1} \to s_1^!\omega^\kappa|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1}.$$

We can associate to these cohomological correspondences Hecke operators which we again write as

$$U_{w,1} \in \mathrm{Hom}(\mathrm{R}\Gamma(X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa), \mathrm{R}\Gamma(X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa)),$$

$$U'_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1, \omega^\kappa), \mathrm{R}\Gamma(X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa)),$$

$$U''_w \in \mathrm{Hom}(\mathrm{R}\Gamma(X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa), \mathrm{R}\Gamma(X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1, \omega^\kappa)).$$

We of course continue to write $U_{w,2} = U'_w \circ U''_w$.

By Lemmas 4.6.3 and 4.6.4, the long exact sequence

$$H^*(X^I_{K,\mathrm{Kli}}(p)_1, \omega^\kappa) \overset{\times \mathrm{Ha}(\mathcal{G}_w)}{\to} H^*(X^I_{K,\mathrm{Kli}}(p)_1, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1}) \to H^*(X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1, \omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p-1})$$

is $U_{w,2}$- and $U_{w,1}$-equivariant.

Lemma 4.6.7. *We have commutative diagrams*

$$\begin{tikzcd} u_2^*\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1} \arrow[r, "U_{w,1}"] \arrow[d, "u_2^*\mathrm{Ha}'(\mathcal{G}_w)"] & u_1^!\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1} \arrow[d, "u_1^*\mathrm{Ha}'(\mathcal{G}_w)"] \\ u_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1}) \arrow[r, "U_{w,1}"] & r_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1}) \end{tikzcd}$$

$$\begin{tikzcd} r_2^*\omega^\kappa|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1} \arrow[r, "U'_w"] \arrow[d, "r_2^*\mathrm{Ha}'(\mathcal{G}_w)"] & r_1^!\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1} \arrow[d, "r_1^*\mathrm{Ha}'(\mathcal{G}_w)"] \\ r_2^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1}) \arrow[r, "U'_w"] & r_1^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1}) \end{tikzcd}$$

$$\begin{tikzcd} s_1^*\omega^\kappa|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1} \arrow[r, "U''_w"] \arrow[d, "s_1^*\mathrm{Ha}'(\mathcal{G}_w)"] & s_2^!\omega^\kappa|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1} \arrow[d, "s_2^*\mathrm{Ha}'(\mathcal{G}_w)"] \\ s_1^*(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K,\mathrm{Kli}}(p)_1}) \arrow[r, "U''_w"] & s_2^!(\omega^\kappa \otimes \det \omega_{\mathcal{G}_w}^{p^2-1}|_{X^{I,=_w 1}_{K'',\mathrm{Kli}}(p)_1}) \end{tikzcd}$$

Proof. See [Pil20, Lem. 10.5.3.1] . □

4.6.8. *Comparison of $U_{w,2}, U_{w,1}$ and $T_w, T_{w,1}$ in a special case.* We fix $J \subset I$. The space $X_{K_p,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1$ carries a finite étale map to the space $X_{K,1}^{I,=_J 1,=_{J^c} 2}$ studied in §4.2.35. This map is given by forgetting the multiplicative groups H_v of order p at the places $v \in J^c$. Therefore, it has degree $(p+1)^{\#J^c}$. Let $\kappa = (k_v, l_v)$ be a classical algebraic weight. We have an injective map

$$H^0(X_{K,1}^{I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D)) \to H^0(X_{K,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D)).$$

We assume that $l_v \geq 3$ if $v \in I^c$, that $k_v \geq 3$, $l_v \geq 2$ if $v \in I$, and moreover that $l_v \geq p+1$ and $k_v \geq 2p+3$ if $v \in J$. On the left hand side, we have an action of T_w for $w|p$ and $T_{w,1}$ for $w \in I^c$. On the right hand side, we have an action of $U_{w,2}$ for $w|p$ and $U_{w,1}$ for $w \in I^c$. This follows from the fact that all these Hecke operators have been proved to commute with the Hasse invariants (by Lemmas 4.2.7, 4.2.15, 4.2.19, 4.6.3, 4.6.4, 4.6.7).

The main result of this subsection is:

Proposition 4.6.9. *There is a universal constant C depending only on p and F but not on the tame level K^p such that if $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w|p$, and $l_w \geq C$ for all $w \in I^c$, then:*

(1) The operator $U^I = \prod_{w|p} U_{w,2} \prod_{w \in I^c} U_{w,1}$ is locally finite on

$$H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D)).$$

(2) Let $\widetilde{T}^I = \prod_{w|p} T_w \prod_{w \in I^c} T_{w,1}$. The map

$$e(\widetilde{T}^I) H^0(X_{K,1}^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D)) \to e(U^I) H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$$

is an isomorphism.

(3) This isomorphism is equivariant for the action of $T_{w,1}$ on the left and $U_{w,1}$ on the right for all $w \in I^c$ and of T_w and $U_{w,2}$ for all $w \in J$.

This result establishes a first relation between the cohomology at Klingen level and spherical level and will allow us to reduce a big proportion of the proof of Theorem 4.6.1 to Theorem 4.2.1.

Remark 4.6.10. One can interpret this result as saying that the ordinarity condition prevents the existence of "newforms" of Klingen level.

We have a finite étale map:

$$X_{K,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1 \to (X_K^{I,=_J 1,=_{J^c} 2})_1$$

which parametrizes multiplicative subgroups of order p, $H_w \subset \mathcal{G}_w[p]$ for all $w \in J^c$. We introduce various Hecke operators that decrease the level at places $w \in J^c$ and compare them with our existing Hecke operators.

We first define a correspondence for each $w \in J^c$ (with X_w defined below),

$$\begin{tikzcd} & X_w \arrow[dl, "x_2"'] \arrow[dr, "x_1"] & \\ (X_K^{I,=_J 1,=_{J^c} 2})_1 & & (X_K^{I,=_J 1,=_{J^c} 2})_1 \end{tikzcd}$$

as follows. We let $x_1 : X_w \to (X_K^{I,=_J 1,=_{J^c} 2})_1$ be the natural forgetful map, where X_w parametrizes subgroups $L_w \subset \mathcal{G}_w[p^2]$, where L_w is totally isotropic of étale rank p^3 and multiplicative rank p. A standard computation shows that x_1

is finite flat. For a suitable choice of polyhedral decomposition, there is a map $x_2 : X_w \to (X_K^{I,=_J 1,=_{J^c} 2})_1$, which on the p-divisible group is given by $\mathcal{G} \mapsto \mathcal{G}/L_w$ (since we are only dealing with H^0 cohomology groups, we will for the most part suppress the discussion of the boundary in this section).

We can define an operator, using the usual procedure, associated to X_w, $\widetilde{T}_w \in \mathrm{End}(H^0((X_K^{I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D)))$.

If we denote by $X_{K,\mathrm{Kli}_w}^{I,=_J 1,=_{J^c} 2}(p)_1 \to (X_K^{I,=_J 1,=_{J^c} 2})_1$ the finite étale cover that parametrizes subgroups $H_w \subset \mathcal{G}_w[p]$ of order p, then we observe that the projection x_2 lifts to a map $X_w \to X_{K,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1$ by sending $(\mathcal{G}, L_w)$ to $(\mathcal{G}/L_w, \mathcal{G}_w[p]/L_w)$ and therefore one can promote $\widetilde{T}_w$ to maps $\widetilde{T}'_w$ and $\widetilde{T}''_w$ fitting in a commutative diagram (where vertical maps are the injections given by the obvious pull back maps):

$$\begin{tikzcd}
H^0(X_{K,\mathrm{Kli}_w}^{I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D)) \arrow[r, "\widetilde{T}''_w"] \arrow[dr, "\widetilde{T}'_w"] & H^0(X_{K,\mathrm{Kli}_w}^{I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D)) \\
H^0((X_K^{I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D)) \arrow[u] \arrow[r, "\widetilde{T}_w"] & H^0((X_K^{I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D)) \arrow[u]
\end{tikzcd} \tag{4.6.11}$$

On the other hand we have already defined a Hecke operator T_w on $H^0((X_K^{I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D))$. We also have a chain of finite étale maps:

$$X_{K,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1 \to X_{K,\mathrm{Kli}_w}^{I,=_J 1,=_{J^c} 2}(p)_1 \to (X_K^{I,=_J 1,=_{J^c} 2})_1$$

where the first map forgets the multiplicative subgroup of order p, $H_{w'} \subset \mathcal{G}_{w'}[p]$ for $w \in J^c \setminus \{w\}$. We have defined an operator $U_{w,2}$ on $H^0(X_{K,\mathrm{Kli}}^{I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$, but clearly it descends to an operator on $H^0(X_{K,\mathrm{Kli}_w}^{I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$ because only the Klingen level structure at w matters in the definition of $U_{w,2}$.

Lemma 4.6.12. *Assume that $k_w - l_w \geq 1$.*

(1) We have $\widetilde{T}_w = T_w$.
(2) We have $U_{w,2} \circ \widetilde{T}''_w = U_{w,2} \circ U_{w,2}$.

Proof. See [Pil20, Lem. 11.1.1.1, Lem. 11.1.1.3]. □

Lemma 4.6.13. *Let $w \in J$. The canonical map $H^0(X_{K,1}^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D)) \to H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$ intertwines the actions of T_w and $U_{w,2}$.*

Proof. By Lemma 4.2.19 we have $T_w = T'_{w,et} \circ T''_{w,et}$, which corresponds to $U_{w,2}$ by definition of the right hand side. □

Lemma 4.6.14. *Let $w \in I^c$. The canonical map $H^0(X_{K,1}^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D)) \to H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$ intertwines the actions of $T_{w,1}$ and $U_{w,1}$.*

Proof. By Lemma 4.2.6, we have $T_{w,1} = T_{w,1}^{et}$, which corresponds to $U_{w,1}$ by definition on the right hand side. □

Corollary 4.6.15. *Suppose that we have $l_w \geq 2$ and $k_w - l_w \geq 1$ for all $w \in J^c$. Then the action of $\prod_{w \in J^c} U_{w,2}$ on $H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p)_1, \omega^\kappa(-D))$ is locally finite,*

the action of $\prod_{w\in J^c}\tilde{T}_w$ *is locally finite on* $H^0(X_{K,1}^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D))$*, and the map*

$$e(\prod_{w\in J^c}\tilde{T}_w)H^0(X_{K,1}^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D))\to e(\prod_{w\in J^c}U_{w,2})H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$$

is surjective.

Proof. Combining the diagrams (4.6.11) for all $w\in J^c$, we see that there is a commutative diagram:

$$\begin{tikzcd}
H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D)) \arrow[r, "\prod_{w\in J^c}\tilde{T}''_w"] \arrow[dr, "\prod_{w\in J^c}\tilde{T}'_w"] & H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D)) \\
H^0(X_{K,1}^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D)) \arrow[u, "\iota"] \arrow[r, "\prod_{w\in J^c}\tilde{T}_w"] & H^0(X_{K,1}^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D)) \arrow[u, "\iota"']
\end{tikzcd}$$

where the vertical maps ι are the natural injections. We deduce from Lemma 4.6.12 that $\prod_{w\in J^c}U_{w,2}\circ(\prod_{w\in J^c}\tilde{T}_w)''=\prod_{w\in J^c}U_{w,2}\circ\prod_{w\in J^c}U_{w,2}$. The argument now follows the proofs of [Pil20, Cor. 11.1.1.1, Cor. 11.1.1.2]. We deduce that $\prod_{w\in J^c}\tilde{T}_w$ acts locally finitely on $H^0(X_{K,1}^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D))$ by Lemma 4.2.15. It follows that for any $f\in H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$, there is a $\prod_{w\in J^c}\tilde{T}_w$-stable finite-dimensional vector space V containing $(\prod_{w\in J^c}\tilde{T}_w)'f$, and then the subspace of $H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$ spanned by $\iota(V)$, $\prod_{w\in J^c}U_{w,2}\iota(V)$, f, and $\prod_{w\in J^c}U_{w,2}f$ is finite-dimensional and $\prod_{w\in J^c}U_{w,2}$-stable, so $\prod_{w\in J^c}U_{w,2}$ acts locally finitely, as claimed.

To prove the claimed surjectivity, if $f\in e(\prod_{w\in J^c}U_{w,2})H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$, then one checks from the definitions that we have

$$f=e(\prod_{w\in J^c}U_{w,2})\iota(e(\prod_{w\in J^c}\tilde{T}_w)(\prod_{w\in J^c}\tilde{T}'_w)(\prod_{w\in J^c}U_{w,2})^{-1}f). \qquad \square$$

It remains to prove the injectivity of the map considered in Proposition 4.6.9 (2). This will be done by exhibiting an inverse up to a certain power of p. For this reason, it is necessary to lift the situation to a $\mathbf{Z}_p$-flat base. This is done by considering certain formal schemes. Let us denote by $\mathfrak{X}_K^{G_1,I,=_J1,=_{J^c}2}$ the formal completion of $\mathfrak{X}_K^{G_1,I,=_{J^c}2}$ along $X_{K,1}^{G_1,I,=_J1,=_{J^c}2}$. We also denote by $\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)$ the formal completion of $\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_{J^c}2}(p)$ along $X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1$. We denote by I the ideal of definition of these formal schemes. Observe that $p\in I$ and that $I/p=(\mathrm{Ha}(\mathcal{G}_w)\det\omega_{\mathcal{G}_w}^{(1-p)},\ w\in J)$.

We consider the modules

$$H^0(\mathfrak{X}_K^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D)) \text{ and } H^0(\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p),\omega^\kappa(-D)),$$

which are I-adically complete and separated, and also $\mathbf{Z}_p$-flat. Moreover, the natural map $H^0(\mathfrak{X}_K^{G_1,I,=_J1,=_{J^c}2},\omega^\kappa(-D))\to H^0(\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p),\omega^\kappa(-D))$ reduces modulo I to the map $H^0((X_K^{G_1,I,=_J1,=_{J^c}2})_1,\omega^\kappa(-D))\to H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$ (we are using here that $(X_K^{G_1,I,=_J1,=_{J^c}2})_1$ and $X_{K,\mathrm{Kli}}^{G_1,I,=_J1,=_{J^c}2}(p)_1$ have affine image in the minimal compactification and thus that higher cuspidal cohomology over these spaces vanishes).

We can lift the map

$$\tilde{T}_w : H^0((X_K^{G_1,I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D)) \to H^0((X_K^{G_1,I,=_J 1,=_{J^c} 2})_1, \omega^\kappa(-D))$$

to a map $\tilde{T}_w : H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D)) \to H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D))$ (the correspondences X_w lift to correspondences on the formal schemes).

There is a trace map

$$\mathrm{Tr} : H^0(\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p), \omega^\kappa(-D)) \to H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D))$$

associated to the finite étale map $\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c} 2}(p) \to \mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c} 2}$.

Lemma 4.6.16. *For any $n \in \mathbf{Z}_{\geq 1}$, we have the congruence*

$$\mathrm{Tr} \circ (\prod_{w \in J^c} U_{w,2})^n \circ \iota(f) \equiv p^{\# J^c} (\prod_{w \in J^c} \tilde{T}_w)^n (f) \pmod{p^{\inf_{w \in J^c} k_w - l_w}}$$

for any $f \in H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c} 2}, \omega^\kappa(-D))$.

Proof. We have

$$\mathrm{Tr} \circ (\prod_{w \in J^c} U_{w,2})^n \circ \iota(f(\mathcal{G}, \omega)) = \frac{1}{p^{n \sum_{w \in J^c} l_w + 3}} \sum_{w \in J^c} \sum_{L_{w,n}} (\sum_{H_w} f(\mathcal{G}/(\oplus_w L_{w,n}), \omega')$$

where H_w runs over all multiplicative subgroups of rank p of $\mathcal{G}_w[p]$ and $L_{w,n}$ runs over all totally isotropic subgroups of order p^{3n} of $\mathcal{G}_w[p^{2n}]$, with trivial intersection with H_w (this implies that $L_{w,n}$ is locally in the étale topology an extension of $\mathbf{Z}/p^n\mathbf{Z} \oplus \mathbf{Z}/p^{n^2}\mathbf{Z}$ by μ_{p^n}), and where ω is a trivialization of $\omega_{\mathcal{G}}$ and ω' is a rational trivialization of $\omega_{\mathcal{G}/(\oplus_w L_{w,n})}$, defined by the condition that $\pi^* \omega' = \omega$ for the isogeny $\mathcal{G} \to \mathcal{G}/(\oplus L_{w,n})$. Given a group $L_{w,n}$, we can find p subgroups H_w of order p and of multiplicative type such that $L_{w,n} \cap H_w = \{0\}$. This means that the groups $L_{w,n}$ in the formula defining $\mathrm{Tr} \circ (\prod_{w \in J^c} U_{w,2}) \circ \iota(f)$ occur with multiplicity $p^{\# J^c}$. On the other hand,

$$\prod_{w \in J^c} \tilde{T}_w f(\mathcal{G}, \omega) = \frac{1}{p^{\sum_{w \in J^c} l_w + 3}} \sum_{w \in J^c} \sum_{L_w} f(\mathcal{G}/(\oplus_w L_w), \omega')$$

where L_w runs over all totally isotropic subgroups of order p^3 of $\mathcal{G}_w[p^2]$ with multiplicative rank 1. Now we observe that

$$(\prod_{w \in J^c} \tilde{T}_w)^n f(\mathcal{G}, \omega) \equiv \frac{1}{p^{n \sum_{w \in J^c} l_w + 3}} \sum_{w \in J^c} \sum_{L_{w,n}} f(\mathcal{G}/(\oplus_w L_{w,n}), \omega') \pmod{p^{\inf_{w \in J^c} k_w - l_w}}$$

where $L_{w,n}$ runs over all totally isotropic subgroups of order p^{3n} of $\mathcal{G}_w[p^{2n}]$, which are locally in the étale topology an extension of $\mathbf{Z}/p^n\mathbf{Z} \oplus \mathbf{Z}/p^{n^2}\mathbf{Z}$ by μ_{p^n}. Indeed, if we write

$$(\prod_{w \in J^c} \tilde{T}_w)^n f(\mathcal{G}, \omega) = \frac{1}{p^{n \sum_{w \in J^c} l_w + 3}} \sum_{w \in J^c} \sum_{L'_{w,n}} f(\mathcal{G}/(\oplus_w L'_{w,n}), \omega')$$

using the definition of $\prod_{w \in J^c} \tilde{T}_w$, we find that all the groups $L_{w,n}$ appear exactly one time among the groups $L'_{w,n}$, and that all the remaining groups precisely contain the multiplicative subgroup of $\mathcal{G}_w[p]$ (and these give a contribution divisible by $p^{(k_w - l_w)}$). $\square$

Lemma 4.6.17. *Assume that for all $w \in J^c$, we have $k_w - l_w > p^{\#J^c}$. Then the natural map*

$$e(\prod_{w\in J^c} \tilde{T}_w)H^0(X_{K,1}^{G_1,I,=_J 1,=_{J^c}2},\omega^\kappa(-D)) \to e(\prod_{w\in J^c} U_{w,2})H^0(X_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c}2}(p)_1,\omega^\kappa(-D))$$

is bijective.

Proof. We will show that the map:

$$\iota : e(\prod_{w\in J^c} \tilde{T}_w)H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c}2},\omega^\kappa(-D)) \to e(\prod_{w\in J^c} U_{w,2})H^0(\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I,=_J 1,=_{J^c}2}(p),\omega^\kappa(-D))$$

is bijective (note that it is legitimate to apply the ordinary projectors on these spaces, because they can be written as projective limits (modding out by I^n) of spaces carrying a locally finite action). The result will then follow by taking reduction modulo I. The map is surjective by Corollary 4.6.15 (and using I-adic approximation). It remains to prove injectivity. Let us take

$$f \in e(\prod_{w\in J^c} \tilde{T}_w)H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c}2},\omega^\kappa(-D)),$$

with $f \neq 0$ and $\iota(f) = 0$. Without loss of generality, we can suppose that

$$f \notin pe(\prod_{w\in J^c} \tilde{T}_w)H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c}2},\omega^\kappa(-D)).$$

It follows from Lemma 4.6.16 that $\mathrm{Tr}(\iota f) \equiv p^{\#J^c} f \pmod{p^{\inf_w k_w - l_w}}$. Therefore, $f \in p^{(\inf_w k_w - l_w)-\#J^c} e(\prod_{w\in J^c} \tilde{T}_w)H^0(\mathfrak{X}_K^{G_1,I,=_J 1,=_{J^c}2},\omega^\kappa(-D))$. This is a contradiction. □

Proof of Proposition 4.6.9. This is immediate from Corollary 4.6.15, Lemma 4.6.17, and Lemmas 4.6.12, 4.6.13 and 4.6.14. □

4.6.18. *Comparison of the cohomology on $X_{K,\mathrm{Kli}}^{I,G_1}(p)_1$ and $X_{K,1}^{I,G_1}$.* We now deduce the following proposition.

Proposition 4.6.19. *There is a universal constant C depending only on p and F but not on the tame level K^p such that if $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w|p$, and $l_w \geq C$ for all $w \in I^c$, then the operator U^I is locally finite on $\mathrm{R}\Gamma(X_{K,\mathrm{Kli}}^{I,G_1}(p)_1,\omega^\kappa(-D))$ and there is a canonical quasi-isomorphism:*

$$e(\widetilde{T}^I)\mathrm{R}\Gamma(X_{K,1}^{I,G_1},\omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(X_{K,\mathrm{Kli}}^{I,G_1}(p)_1,\omega^\kappa(-D)).$$

Proof. In §4.2.30, we constructed a complex $K^\bullet$ computing explicitly the cohomology $\mathrm{R}\Gamma(X_{K,1}^{I,G_1},\omega^\kappa(-D))$. We recall that $K^k =$

$$\oplus_{J\subset I,\#J=k} \varinjlim_{\times\prod_{w\in J}\mathrm{Ha}(\mathcal{G}_w)} H^0\big(X_{K,1}^{G_1,I,\geq_{J^c}2},\omega^\kappa(-D)\otimes\bigotimes_{w\in J}(\det\mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n))\big).$$

In exactly the same way, there is a complex $L^\bullet$ computing $\mathrm{R}\Gamma(X_{K,\mathrm{Kli}}^{I,G_1}(p)_1,\omega^\kappa(-D))$, such that $L^k =$

$$\oplus_{J\subset I,\#J=k} \varinjlim_{\times\prod_{w\in J}\mathrm{Ha}(\mathcal{G}_w)} H^0\big(X_{K,\mathrm{Kli}}^{G_1,I,\geq_{J^c}2}(p)_1,\omega^\kappa(-D)\otimes\bigotimes_{w\in J}(\det\mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n))\big).$$

It therefore suffices to prove that for each $J \subset I$ and each $n \geq 1$, U^I is locally finite on

$$H^0\big(X^{G_1,I,\geq_{J^c}2}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n))\big),$$

and the map

$$e(\widetilde{T}^I)H^0\big(X^{G_1,I,\geq_{J^c}2}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n))\big) \to$$
$$e(U^I)H^0\big(X^{G_1,I,\geq_{J^c}2}_{K,\mathrm{Kli}}(p)_1, \omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n))\big)$$

is an isomorphism. In the case $n = 1$, this is Proposition 4.6.9, and the general case follows by induction on n, using the short exact sequence

$$0 \to \omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{(n-1)(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^{n-1})) \to$$
$$\omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w)^n)) \to$$
$$\omega^\kappa(-D) \otimes \bigotimes_{w\in J}(\det \mathcal{G}_w)^{n(p-1)}/(\sum_{w\in J}(\mathrm{Ha}(\mathcal{G}_w))) \to 0$$

and the acyclicity of these sheaves (for which see the proof of Proposition 4.2.33). □

4.6.20. *The proof of Theorem 4.6.1.*

Lemma 4.6.21. *If $\kappa = ((k_v, l_v)_{v|p})$ is a classical algebraic weight with $l_v = 2$ when $v \in I$ and $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$, then for each $n \geq 2$ there is a diagonal map making a commutative diagram:*

$$\begin{tikzcd}
\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{\mathrm{Kli}}(p^n), \omega^\kappa(-D)) \arrow[r, "U^I"] \arrow[dr, "U^I"] & \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{\mathrm{Kli}}(p^n), \omega^\kappa(-D)) \\
\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{\mathrm{Kli}}(p^{n-1}), \omega^\kappa(-D)) \arrow[r, "U^I"] \arrow[u] & \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{\mathrm{Kli}}(p^{n-1}), \omega^\kappa(-D)) \arrow[u]
\end{tikzcd}$$

Proof. This is an easy computation in the Hecke algebra, see the proof of [Pil20, Thm. 11.3.1]. □

We now make repeated use of Nakayama's lemma for complexes, in the form of [Pil20, Prop. 2.2.1, Prop. 2.2.2]. In fact, we need the following slight strengthening of [Pil20, Prop. 2.2.1], which is proved in the same way; for ease of reference we explain how it follows from results in the literature.

Lemma 4.6.22. *Let R be a complete local Noetherian ring with maximal ideal $\mathfrak{m}$, and let $M^\bullet$ be a bounded complex of $\mathfrak{m}$-adically complete and separated, flat R-modules, with the property that the cohomology groups of $M^\bullet \otimes_R R/\mathfrak{m}$ are finite-dimensional and concentrated in degrees $[a, b]$. Then $M^\bullet$ is a perfect complex, concentrated in degrees $[a, b]$.*

Proof. It follows from [Pil20, Prop. 2.2.1] that $M^\bullet$ is a perfect complex, and it then follows from [KT17, Lem. 2.3, Cor. 2.7] that it is concentrated in degrees $[a, b]$. □

All the complexes we consider below can be represented by bounded complexes of flat, complete and separated $\mathbf{Z}_p$-modules (resp. Λ_I-modules), as can be seen by considering a Čech complex for any finite affine cover, so the hypotheses of Lemma 4.6.22 apply in our situation.

Lemma 4.6.23. *For all classical algebraic weights $\kappa = ((k_v, l_v)_{v|p})$ with $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w|p$, and $l_w \geq C$ for all $w \in I^c$, the operator U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$ and there is a canonical quasi-isomorphism:*

$$e(\widetilde{T}^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_K, \omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D)).$$

Proof. By Proposition 4.6.19, together with [Pil20, Prop. 2.2.2, Prop. 2.3.1], the action of U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p), \omega^\kappa(-D))$, and the map

$$e(\widetilde{T}^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_K, \omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p), \omega^\kappa(-D))$$

is a quasi-isomorphism. It follows easily from Lemma 4.6.21 that U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$ and that the map $e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p), \omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{I,G_1}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$ is a quasi-isomorphism, as required. □

Let $\kappa = (k_v, l_v)$ be a classical algebraic weight. Let $K\omega^\kappa$ denote the kernel of the surjection of Corollary 4.3.9, so that over $\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty)$, so we have a short exact sequence of sheaves

$$0 \to K\omega^\kappa \to \omega^\kappa \to \Omega^\kappa \to 0.$$

A key step in the comparison between the ordinary forms of these weights is the following basic lemma.

Lemma 4.6.24. *For any $w|p$, we have $U_{w,2} \in p\mathrm{End}(\mathrm{R}\Gamma(\mathfrak{X}^I_{K,\mathrm{Kli}}(p^\infty), K\omega^\kappa))$.*

Proof. This follows immediately from an examination of (4.5.12). □

Lemma 4.6.25. *For all classical algebraic weights $\kappa = ((k_v, l_v)_{v|p})$ with $l_v = 2$ when $v \in I$ and $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$, the operator U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$ and $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^\kappa(-D))$, and the map*

$$e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p), \omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^\kappa(-D))$$

is a quasi-isomorphism.

Proof. We consider the exact triangle

$$\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), K\omega^\kappa(-D)) \to \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$$
$$\to \mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^\kappa(-D)).$$

By Lemma 4.6.24, the operator U^I is topologically nilpotent on

$$\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), K\omega^\kappa(-D)),$$

so in particular it acts locally finitely with $e(U^I) = 0$.

If we further assume that $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w|p$, and $l_w \geq C$ for all $w \in I^c$, then it follows from Lemma 4.6.23 that U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \omega^\kappa(-D))$. Therefore in this case, it follows from the above exact triangle that U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^\kappa(-D))$.

Again using [Pil20, Prop. 2.3.1], we deduce that U^I is locally finite on the complex $\mathrm{R}\Gamma(\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}}(p^\infty), \Omega^\kappa(-D))$ for any weight κ, and therefore (again using the

above exact triangle) it is also locally finite on $\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{G_1,I}(p^\infty),\omega^\kappa(-D))$ for any weight κ, as required. □

Lemma 4.6.26. *For all classical algebraic weights $\kappa = ((k_v,l_v)_{v|p})$ with $l_v = 2$ when $v \in I$, $k_v \equiv l_v \equiv 2 \pmod{p-1}$, for all $v|p$, $l_w \geq 2$ for all w, $k_w - l_w \geq C$ for all $w \in I$, and $l_w \geq C$ for all $w \in I^c$, the complex $e(U^I)\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{I,G_1}(p^\infty),\Omega^\kappa(-D))$ is a perfect complex of $\mathbf{Z}_p$-modules concentrated in degrees $[0,\#I]$.*

Proof. This follows from Lemma 4.6.25, Lemma 4.6.23 and Theorem 4.2.1 (noting that T^I divides $\widetilde{T}^I$, so that $e(\widetilde{T}^I)\mathrm{R}\Gamma(X_1^{G_1,I},\omega^\kappa(-D))$ is a direct summand of $e(T^I)\mathrm{R}\Gamma(X_1^{G_1,I},\omega^\kappa(-D)))$. □

Lemma 4.6.27. *The operator U^I is locally finite on $\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{I,G_1}(p^\infty),\Omega^{\kappa_I}(-D))$, and $e(U^I)\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{I,G_1}(p^\infty),\Omega^{\kappa_I}(-D))$ is a perfect complex of Λ_I-modules concentrated in degree $[0,\#I]$.*

Proof. This follows from Lemma 4.6.26 by Nakayama's lemma, in the form of Lemma 4.6.22 and [Pil20, Prop. 2.3.1]. □

Proof of Theorem 4.6.1. Parts (1) and (2) are Lemma 4.6.27. Part (3) is Lemma 4.6.25, together with Lemma 4.6.21, which shows that the natural map

$$e(U^I)\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{I,G_1}(p),\omega^\kappa(-D)) \to e(U^I)\mathrm{R}\Gamma(\mathfrak{X}_{K,\mathrm{Kli}}^{I,G_1}(p^\infty),\omega^\kappa(-D))$$

is a quasi-isomorphism. Part (4) follows from Theorem 4.2.1, together with Proposition 4.6.19 and Lemma 4.6.25. □

5. Doubling

In this section, we prove a doubling result (see Theorems 5.8.6 and 5.8.4) which is the key ingredient for proving local–global compatibility in §7.9. The general ideal of doubling is that certain spaces of ordinary low weight modular forms admit (at least) two degeneracy maps to spaces of ordinary modular forms of either higher weight or higher level. For example, the space of weight one elliptic modular forms modulo p of level $\Gamma_1(N)$, $p \nmid N$, admits degeneracy maps $f \mapsto \mathrm{Ha}\cdot f$ and $f \mapsto f^p$ (where Ha is the Hasse invariant) to spaces of forms of weight p and level $\Gamma_1(N)$. (Alternatively, after dividing by Ha, these degeneracy maps can also be thought of as maps from classical forms of level $\Gamma_1(N)$ and weight one to ordinary p-adic modular forms of level $\Gamma_1(N)$ and weight one.) If one can show that the direct sum of two copies of the original space embeds under the direct sum of these degeneracy maps, then, following ideas going back to Gross [Gro90] and isolated and expanded by [Wie14] (see also [CG18] for further exploitation of these ideas), one can make deductions about the local properties of the Galois representations of interest.

Let us explicate this in the example of weight one forms mentioned above (the following is implicit in the first few lines of [Gro90, p.499] and explicit in [Edi92, Prop. 2.7]). If f is a weight one elliptic modular cuspidal eigenform with Nebentypus character χ and T_p-eigenvalue a_p satisfying $a_p^2 \neq 4\chi(p)$, one can show that the associated Galois representation is unramified at p in the following way. Since the polynomial $X^2 - a_pX + \chi(p)$ has distinct roots, one can show using the degeneracy maps above (and having established doubling) that there are two weight p ordinary forms congruent to f with level $\Gamma_1(N)$ and T_p-eigenvalues given by the roots α and β of $X^2 - a_pX + \chi(p)$. Using the known properties of the corresponding Galois

representations, one shows that the restriction to p is an extension of distinct unramified characters, and thus that the extension is split (because the representations corresponding to the two weight p forms are extensions in the opposite orders).

The above argument for local–global compatibility at p works equally well in the ordinary symplectic case once we have established a doubling theorem, and we will use this in §7.9 below. Before proceeding, we begin by recalling the doubling argument in more detail in the case of $\mathrm{GL}_2/\mathbf{Q}$.

5.1. **The case of** $\mathrm{GL}_2/\mathbf{Q}$**.** For the moment, let X denote the special fibre of a classical modular curve of level $\Gamma_1(N)$ with $N \geq 5$ with $p \nmid N$, and let ω denote the usual invertible line bundle on X (as in [Gro90, §2]). The doubling strategy of [CG18, Cal18] may be reduced to ruling out the existence of simultaneous eigenforms $f \in H^0(X, \omega)$ for the operators T_p and U_p. This is easily seen: indeed if f is a simultaneous eigenform for T_p and U_p, then it is also an eigenform for $V_p = T_p - U_p$, which is immediately seen to be impossible by examining the action on q-expansions. This argument does not directly generalize to the symplectic case (even over $\mathbf{Q}$), and instead, the paper [CG20] employs a rather labyrinthian argument involving q-expansions to prove an analogous result for $\mathrm{GSp}_4/\mathbf{Q}$. In this paper, we give a different argument which is based on analyzing the behavior of the U_p operator at the non ordinary locus. This argument in this form appears to be new even for modular forms of weight one (although there are certainly some echos of this argument in papers such as [Joc82, Ser73, Cai14]), and so we present it first as a warm up for the general symplectic case.

If $f \in H^0(X, \omega)$, we may think of $U_p f$ as a section of $H^0(X \setminus SS, \omega)$ for the finite set SS of supersingular points of X. We claim that there is a commutative diagram

$$\begin{tikzcd} H^0(X,\omega) \arrow[r, "\mathrm{Ha}\cdot U_p"] \arrow[d] & H^0(X,\omega^p) \arrow[d] \\ H^0(SS,\omega) \arrow[r] & H^0(SS,\omega^p) \end{tikzcd} \tag{5.1.1}$$

where the vertical maps are the natural restriction maps, and the lower horizontal map is an isomorphism. The existence of such a diagram can be proved in several ways; for example it can be checked in the same way as the corresponding statements for GSp_4/F later in this section, by using the Kodaira–Spencer isomorphism to describe the U_p operator as a trace map on differentials.

Suppose that f is a U_p-eigenform in $H^0(X, \omega)$ with non-zero eigenvalue. Considering the commutative diagram (5.1.1), we see that since $\mathrm{Ha} \cdot U_p f$ maps to zero in $H^0(SS, \omega^p)$, the restriction of f to SS must vanish. Thus $f = \mathrm{Ha} \cdot g$ for some $g \in H^0(X, \omega^{2-p})$, and this cohomology group vanishes if $p > 2$, so $f = 0$ in this case. If $p = 2$, the only non-zero sections of $H^0(X, \mathcal{O}_X)$ are constants, and we deduce that f is a multiple of the Hasse invariant.

In the rest of this section we prove a generalization of this to the Hilbert–Siegel case. The analogue of the commutative diagram (5.1.1) in the Siegel case (with $F = \mathbf{Q}$) is the following commutative diagram (where we write $Y^{\geq 1}$ for the locus in the interior of the special fibre of the Shimura variety with Klingen level H which is multiplicative, and we write $Y^{=1}$ for the divisor where the abelian variety is non ordinary.)

$$\begin{tikzcd}
H^0(Y^{\geq 1}, \omega^2) \arrow[r, "\mathrm{Ha}\cdot U_{\mathrm{Iw}(p),1}"] \arrow[d] & H^0(Y^{\geq 1}, \omega^{p+1}) \arrow[d] \\
H^0(Y^{=1}, \omega^2) \arrow[r] & H^0(Y^{=1}, \omega^{p+1})
\end{tikzcd}$$

However, in contrast to the modular curve case, the map on the bottom line of this diagram is probably not injective, so we cannot conclude as before. Instead, we construct a larger commutative diagram

$$\begin{tikzcd}
H^0(Y^{\geq 1}, \omega^2) \arrow[r, "\mathrm{Ha}\cdot U_{\mathrm{Iw}(p),1}"] \arrow[rr, bend left, "U_{p,2}"] \arrow[d] & H^0(Y^{\geq 1}, \omega^{p+1}) \arrow[d] & H^0(Y^{\geq 1}, \omega^2) \arrow[d] \\
H^0(Y^{=1}, \omega^2) \arrow[r] & H^0(Y^{=1}, \omega^{p+1}) \arrow[r] & H^0(Y^{=1}, \omega^2).
\end{tikzcd}$$

If we assume that $f \in H^0(Y^{\geq 1}, \omega^2)$ is also a $U_{p,2}$-eigenform with nonzero eigenvalue (which suffices for our purposes), we can use this diagram to make a similar argument to the above, considering the composite morphisms from the top left to the lower right hand corner. It may help the reader to note that there is an analogous diagram for GL_2:

$$\begin{tikzcd}
H^0(X, \omega) \arrow[r, "\mathrm{Ha}\cdot U_p"] \arrow[rr, bend left, "\langle p \rangle"] \arrow[d] & H^0(X, \omega^p) \arrow[d] & H^0(X, \omega) \arrow[d] \\
H^0(SS, \omega) \arrow[r] & H^0(SS, \omega^p) \arrow[r] & H^0(SS, \omega)
\end{tikzcd}$$

(again, the existence of this diagram can be checked in the same way as our calculations below). We see that if $U_p f$ has no poles, then the image of f in the bottom right hand copy of $H^0(SS, \omega)$ vanishes; since the diamond operator $\langle p \rangle$ is an isomorphism, it follows that the restriction of f to SS vanishes, and we conclude as before.

There is an additional complication in the Hilbert–Siegel case, which is that rather than considering the entire Shimura variety, we are only working on an open subspace $X^{\geq_{v \in I} 1, \geq_{v \in I^c} 2}_{K_p(I)K^p,1}$. This means that the vanishing of the space of (partial) negative weight modular forms is not obvious. We sketch a proof for this vanishing in §5.9 below, using Fourier–Jacobi expansions (which ultimately reduces to the vanishing of spaces of Hilbert modular forms of partial negative weight), but we do not rely on this result. Instead, we give a complete proof of a slightly weaker result which is nonetheless sufficient for our purposes; this argument does not use the

boundary, but rather considers the behavior of another Hecke operator $U_{\mathrm{Kli}(w),1} - U_{\mathrm{Iw}(w),1}$ (called Z_w below) along the w-non-ordinary locus.

5.2. **Conventions.** Throughout this section, we fix a set $I \subset S_p$ and a prime $w \in I$. Recall that, as in §4.3.4, for each subset $I \subset S_p$ we set

$$K_p(I) = \prod_{v \in I} \mathrm{Kli}(v) \prod_{v \in I^c} \mathrm{Iw}(v),$$

and we write $X_1^I := X_{K_p(I)K^p,1}^{\geq_{v\in I}1,\geq_{v\in I^c}2}$. We will use the following simplified notation:

- We write X_1^I for the space $X_{K_p(I)K^p,1}^{\geq_{v\in I}1,\geq_{v\in I^c}2}$.
- We write $X_1^{I,=_w2}$ for the open subspace where $A[w^\infty]$ is ordinary.
- We write $X_1^{I,=_w1}$ for the (reduced) complement of this open subspace, which is a divisor in X_1^I.

We also write Y_1^I, $Y_1^{I,=_w2}$ and $Y_1^{I,=_w1}$ for their interiors. We use the analogous notation $\mathfrak{X}^I$, $\mathfrak{Y}^I$ etc. for the corresponding formal schemes. We denote $\det \omega_{\mathcal{G}}$ by ω and $\det \omega_{\mathcal{G}_w}$ by ω_w. We will finally denote the partial Hasse invariant $\mathrm{Ha}(\mathcal{G}_w) \in H^0(X_1^I, \omega_w^{p-1})$ by Ha_w.

5.3. **The operator $U_{\mathrm{Kli}(w),1}$.** We now define a Hecke operator $U_{\mathrm{Kli}(w),1}$ (see also § 4.5.6). We again consider the p-adic completion of the correspondence considered in §3.9.17

$$\begin{tikzcd} & \mathfrak{X}_{K'} \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ \mathfrak{X}_K & & \mathfrak{X}_K \end{tikzcd}$$

where we recall that $K = K^pK_p$ with $K_p = \prod_v \mathrm{GSp}_4(\mathcal{O}_{F_v})$ and $K' = K^pK'_p$ with $K'_p = \mathrm{Si}(w) \times \prod_{v\neq w} \mathrm{GSp}_4(\mathcal{O}_{F_v})$.

We can form the fibre product $\mathfrak{X}_{K'} \times_{p_1,\mathfrak{X}_K} \mathfrak{X}^I$. As $\mathfrak{X}^I \to \mathfrak{X}_K$ is étale by Proposition 4.3.3, this inherits the properties of $\mathfrak{X}_{K'}$ deduced from the theories of local models and toroidal compactifications. In particular it is flat over $\mathbf{Z}_p$, normal, Cohen–Macaulay, and the ordinary locus is dense in the special fibre, see §3.4, Theorem 3.5.1, and §4.1. We denote by $\mathfrak{C}_{\mathrm{Kli}(w),1}$ the open and closed formal subscheme of this fibre product where the kernel of the canonical isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ has trivial intersection with the multiplicative group $p_1^*H_w$.

We obtain a correspondence

$$\begin{tikzcd} & \mathfrak{C}_{\mathrm{Kli}(w),1} \arrow[dl, "v_2"'] \arrow[dr, "v_1"] & \\ \mathfrak{X}^I & & \mathfrak{X}^I \end{tikzcd}$$

where $v_1 : \mathfrak{C}_{\mathrm{Kli}(w),1} \to \mathfrak{X}^I$ is induced by the projection $\mathfrak{X}_{K'} \times_{p_1,\mathfrak{X}_K} \mathfrak{X}^I \to \mathfrak{X}^I$ and v_2 is defined as follows: the projection $\mathfrak{X}_{K'} \times_{p_1,\mathfrak{X}_K} \mathfrak{X}^I \to \mathfrak{X}_{K'}$ composed with $p_2 : \mathfrak{X}_{K'} \to \mathfrak{X}_K$ induces a map $\mathfrak{C}_{\mathrm{Kli}(w),1} \to \mathfrak{X}_K$ which we would like to lift to a map $v_2 : \mathfrak{C}_{\mathrm{Kli}(w),1} \to \mathfrak{X}^I$. In other words, given a point of $\mathfrak{C}_{\mathrm{Kli}(w),1}$, we need to give multiplicative subgroups of order p, $H'_{w'} \subseteq p_2^*\mathcal{G}_{w'}$ for all $w' \in S_p$. But for all $w' \in S_p$, the kernel of the isogeny $p_1^*\mathcal{G}_{w'} \to p_2^*\mathcal{G}_{w'}$ has trivial intersection with $p_1^*H_{w'}$, (for $w' \neq w$ it is an isomorphism, and for $w' = w$ this was assumed in the definition of $\mathfrak{C}_{\mathrm{Kli}(w),1}$) and we take $H'_{w'}$ to be the image of $p_1^*H_{w'}$ under this isogeny.

Lemma 5.3.1. *We have* $R(v_1)_*\mathcal{O}_{\mathfrak{C}_{\mathrm{Kli}(w),1}} = (v_1)_*\mathcal{O}_{\mathfrak{C}_{\mathrm{Kli}(w),1}}$ *and there is a trace map* $R(v_1)_*\mathcal{O}_{\mathfrak{C}_{\mathrm{Kli}(w),1}} \to \mathcal{O}_{\mathfrak{X}^I}$.

Proof. We have $\mathfrak{X}_K = \mathfrak{X}_{K,\Sigma}$ for a smooth polyhedral cone decomposition Σ and we have $\mathfrak{X}_{K'} = \mathfrak{X}_{K',\Sigma'}$. We can now assume that $\Sigma' = \Sigma$ because we have $R\pi_*\mathcal{O}_{\mathfrak{X}_{K',\Sigma'}} = \mathcal{O}_{\mathfrak{X}_{K',\Sigma}}$ for $\pi : \mathfrak{X}_{K',\Sigma'} \to \mathfrak{X}_{K',\Sigma}$ the projection (we note that the cone decomposition at level K' may not be smooth but we will not need this). Since $\Sigma = \Sigma'$, the map v_1 is quasi-finite, and (since it is proper) is therefore finite. Hence we have $R(v_1)_*\mathcal{O}_{\mathfrak{C}_{\mathrm{Kli}(w),1}} = (v_1)_*\mathcal{O}_{\mathfrak{C}_{\mathrm{Kli}(w),1}}$. Moreover, as $\mathfrak{C}_{\mathrm{Kli}(w),1}$ is Cohen–Macaulay and $\mathfrak{X}^I$ is regular, we deduce that the map v_1 is also flat, and so it has an associated trace map. □

We let $\mathfrak{C}^{=_w 2}_{\mathrm{Kli}(w),1}$ be the open formal subscheme where $v_1^*\mathcal{G}_w$ (or equivalently $v_2^*\mathcal{G}_w$) is ordinary. It restricts to a correspondence over $\mathfrak{X}^{I,=_w 2}$. Over $\mathfrak{C}^{=_w 2}_{\mathrm{Kli}(w),1}$, the multiplicative rank of $\ker(v_1^*\mathcal{G}_w \to v_2^*\mathcal{G}_w)$ is either 0 or 1 (it cannot be 2 because H_w has trivial intersection with $\ker(v_1^*\mathcal{G} \to v_2^*\mathcal{G})$), and hence we have a decomposition

$$\mathfrak{C}^{=_w 2}_{\mathrm{Kli}(w),1} = \mathfrak{C}^{=_w 2,et}_{\mathrm{Kli}(w),1} \coprod \mathfrak{C}^{=_w 2,m\text{-}et}_{\mathrm{Kli}(w),1}$$

where $\mathfrak{C}^{=_w 2,et}_{\mathrm{Kli}(w),1}$ is the locus where the isogeny $v_1^*\mathcal{G} \to v_2^*\mathcal{G}$ is étale, while $\mathfrak{C}^{=_w 2,m\text{-}et}_{\mathrm{Kli}(w),1}$ is the locus where $\ker(v_1^*\mathcal{G} \to v_2^*\mathcal{G})$ has multiplicative rank 1.

For any weight $\kappa = (k_v, l_v)$, we have a map

$$v_2^*\omega^\kappa \to v_1^*\omega^\kappa[1/p]$$

induced from the universal isogeny (we note that we are not assuming $l_w \geq 0$.) Tensoring this map with the trace map of Lemma 5.3.1, we obtain a map of sheaves over $\mathfrak{X}^I$:

$$\Theta_\kappa : (Rv_1)_*v_2^*\omega^\kappa \to \omega^\kappa[1/p].$$

We now define $U_{\mathrm{Kli}(w),1} = p^{-l_w-1}\Theta_\kappa$ if $l_w \leq 2$ and $U_{\mathrm{Kli}(w),1} = p^{-3}\Theta_\kappa$ if $l_w \geq 2$.

Lemma 5.3.2. *We have* $U_{\mathrm{Kli}(w),1} : (Rv_1)_*v_2^*\omega^\kappa \to \omega^\kappa$.

Proof. We follow the same strategy as the proof of Lemma 3.9.18. Both the source and target of this map are locally free sheaves over the smooth formal scheme $\mathfrak{X}^I$. To prove that the map is indeed p-integral, it is enough to prove it over the ordinary locus, and we check it separately on each type of component.

On the component of the map corresponding to $\mathfrak{C}^{=_w 2,et}_{\mathrm{Kli}(w),1}$, the isogeny is étale over the ordinary locus and therefore the map $v_2^*\omega^\kappa \to v_1^*\omega^\kappa[1/p]$ is actually an isomorphism $v_2^*\omega^\kappa \to v_1^*\omega^\kappa$, while the trace map is divisible by p^3 (see the proof of Lemma 3.9.18).

On the component of the map corresponding to $\mathfrak{C}^{=_w 2,m\text{-}et}_{\mathrm{Kli}(w),1}$, the isogeny has multiplicative rank one over the ordinary locus and therefore the map

$$v_2^*\omega^\kappa \to v_1^*\omega^\kappa[1/p]$$

is actually a map

$$v_2^*\omega^\kappa \to p^{l_w}v_1^*\omega^\kappa,$$

while the trace map is divisible by p (again see the proof of Lemma 3.9.18).

Therefore, on the étale component, the map is divisible by p^3 and on the multiplicative-étale component it is divisible by p^{l_w+1}. □

5.4. **The operators $U_{\mathrm{Iw}(w),1}$ and Z_w.** Now we consider some Hecke operators on $\mathfrak{X}^{I,=_w 2}$. The restriction of Θ_κ to $\mathfrak{X}^{I,=_w 2}$ decomposes as a sum $\Theta_\kappa = \Theta_\kappa^{et} + \Theta_\kappa^{m\text{-}et}$, according to the decomposition of $\mathfrak{C}^{=_w 2}_{\mathrm{Kli}(w),1}$. We define normalized cohomological correspondences $U_{\mathrm{Iw}(w),1} = p^{-3}\Theta_\kappa^{et}$ and $Z_w = p^{-l_w-1}\Theta_\kappa^{m\text{-}et}$.

Lemma 5.4.1. *The cohomological correspondences $U_{\mathrm{Iw}(w),1}$ and Z_w are p-integral.*

Proof. This follows from the proof of Lemma 5.3.2. □

We have the following identities of Hecke operators over the ordinary locus at w:

(1) $U_{\mathrm{Kli}(w),1} = U_{\mathrm{Iw}(w),1} + p^{l_w-2}Z_w$ if $l_w \geq 2$,
(2) $U_{\mathrm{Kli}(w),1} = p^{2-l_w}U_{\mathrm{Iw}(w),1} + Z_w$ if $l_w \leq 2$.

It follows in particular that:

(1) $U_{\mathrm{Kli}(w),1} = U_{\mathrm{Iw}(w),1} \mod p$ if $l_w > 2$,
(2) $U_{\mathrm{Kli}(w),1} = Z_w \mod p$ if $l_w < 2$.

Another important property is the following:

Proposition 5.4.2. *For any weight κ, we have the following identities of cohomological correspondences over $X_1^{I,=_w 2}$:*

(1) $Z_w \mathrm{Ha}_w = \mathrm{Ha}_w Z_w$,
(2) $U_{\mathrm{Iw}(w),1}\mathrm{Ha}_w = \mathrm{Ha}_w U_{\mathrm{Iw}(w),1}$.

Proof. The correspondence Z_w is the tensor product of the fundamental class (deduced from the trace map normalized by p^{-1}) and a map $v_2^*\omega^\kappa \to v_1^*\omega^\kappa$ which is obtained by normalizing the natural map by a factor p^{-l_w}. It suffices to check that for $\omega^\kappa = \omega_w^{p-1}$ this normalized map matches the Hasse invariants $v_2^*\mathrm{Ha}(\mathcal{G}_w)$ and $v_1^*\mathrm{Ha}(\mathcal{G}_w)$. This is the content of [Pil20, Lem. 6.2.4.1]. The case of $U_{\mathrm{Iw}(w),1}$ is clear because the universal isogeny is étale. □

Finally, we will need the following property:

Proposition 5.4.3. *If $l_w \leq 0$, the cohomological correspondence over X_1^I:*

$$U_{\mathrm{Kli}(w),1} : (v_1)_* v_2^* \omega^\kappa \to \omega^\kappa$$

factors through

$$U_{\mathrm{Kli}(w),1} : (v_1)_* v_2^* \omega^\kappa \to \omega^\kappa(-X_1^{I,=_w 1}).$$

Before giving the proof we need some preparations. Let $BT/\mathbf{F}_p$ be the smooth algebraic stack of quasi-polarized 1-truncated Barsotti–Tate groups of height 2 and dimension 1 over $\operatorname{Spec}\mathbf{F}_p$. Let $Y/\mathbf{F}_p$ be a modular curve of level prime to p. The map $Y \to BT$ is a presentation of BT (that is, it is a smooth surjection). We denote by E the universal object on BT. We have a Cartier divisor $\omega_E^{1-p} \overset{\mathrm{Ha}(E)}{\to} \mathcal{O}_{BT}$ whose support is the non-ordinary locus of BT. Let $\pi : BT_{\mathrm{Iw}} \to BT$ be the representable finite flat map which parametrizes a subgroup $H \subset E$ of order p. Let $Y_0(p)$ be a modular curve of Iwahori level at p. The map $Y_0(p) \to BT_{\mathrm{Iw}}$ is a presentation of BT_{Iw}. Over BT_{Iw} we have a universal morphism $g : E/H \to E$ with kernel H^D (using the polarization to identify E and E^D, E/H and $(E/H)^D$). By differentiating, we get a map of line bundles $dg : \omega_E \otimes \omega_{E/H}^{-1} \to \mathcal{O}_{BT_{\mathrm{Iw}}}$.

Lemma 5.4.4. *We have a canonical factorization $(dg)^{\otimes 2} : (\omega_E \otimes \omega_{E/H}^{-1})^{\otimes 2} \to \pi^*\omega_E^{1-p} \to \mathcal{O}_{BT_{\mathrm{Iw}}}$.*

Proof. It suffices to prove the claim over any presentation of BT_{Iw}. We therefore reduce to proving the statement over the modular curve $Y_0(p)$. The vanishing locus of $\pi^*(\mathrm{Ha}(E))$ is a product of Artinian local rings of length $p+1$ (the degree of π) indexed by the supersingular points. The vanishing locus of dg is the entire irreducible component of $Y_0(p)$ which is degree p over Y via π (this is the component where H is generically étale). Therefore, for any supersingular point $x \in Y$, the image of dg in $\mathcal{O}_{Y_0(p)} \otimes_{\mathcal{O}_Y} k(x)$ defines a closed subscheme of length p, and hence the ideal generated by the image of dg in $\mathcal{O}_{Y_0(p)} \otimes_{\mathcal{O}_Y} k(x)$ is both nilpotent and length 1. It follows that $(dg)^2$ maps to zero in $\mathcal{O}_{Y_0(p)} \otimes_{\mathcal{O}_Y} k(x)$. □

Proof of Proposition 5.4.3. Let κ be a weight with $l_w \leq 2$. Let κ' be another weight with $(k_v, l_v) = (k'_v, l'_v)$ for $v \neq w$, $k_w - l_w = k'_w - l'_w$ and $l'_w = 2$. Let us denote by $U_{\mathrm{Kli}(w),1}(2) : (v_1)_* v_1^* \omega^{\kappa'} \to \omega^{\kappa'}$ the cohomological correspondence in weight κ'. Let $U_{\mathrm{Kli}(w),1} : (v_1)_* v_1^* \omega^{\kappa} \to \omega^{\kappa}$ be the cohomological correspondence in weight κ. The proof of Lemma 5.3.2 shows that the map $v_2^* \det \omega_w^{l_w-2} \to v_1^* \det \omega_w^{l_w-2}[1/p]$ induces a regular map $p^{l_w-2} v_2^* \det \omega_w^{l_w-2} \to v_1^* \det \omega_w^{l_w-2}$, and that moreover $U_{\mathrm{Kli}(w),1}$ is obtained from $U_{\mathrm{Kli}(w),1}(2)$ by twisting by this map. It thus suffices to show that on the special fibre, $p^{l_w-2} v_2^* \det \omega_w^{l_w-2} \to v_1^* \det \omega_w^{l_w-2}$ factors through $v_1^* \det \omega_w^{l_w-2}(-X_1^{I,=1})$ when $l_w \leq 0$.

This statement is local in a neighbourhood of $X_1^{I,=1}$ and we can therefore replace $\mathfrak{X}^I$ by its completion along this closed subscheme. We may also work on the interior of the moduli space, as the interior of the divisor $X_1^{I,=1}$ is dense. Therefore, we may suppose that $\mathcal{G}_w$ comes equipped with a multiplicative sub-Barsotti–Tate subgroup $\mathcal{G}_w^m$ of rank 1, and we denote by $\mathcal{G}_w^{oo} = (\mathcal{G}_w^m)^\perp / \mathcal{G}_w^m$, which is a Barsotti–Tate group scheme of height 2 and dimension 1. The isogeny $v_1^* \mathcal{G}_w \to v_2^* \mathcal{G}_w$ induces an isomorphism $v_1^* \mathcal{G}_w^m \to v_2^* \mathcal{G}_w^m$ and a degree p map $v_1^* \mathcal{G}_w^{oo} \to v_2^* \mathcal{G}_w^{oo}$.

The normalized map $p^{-1} v_2^* \omega_w^{-1} \to v_1^* \omega_w^{-1}$ is the tensor product of the isomorphism $v_2^* \omega_{\mathcal{G}_w^m}^{-1} \to v_1^* \omega_{\mathcal{G}_w^m}^{-1}$ and the map: $p^{-1} v_2^* \omega_{\mathcal{G}_w^{oo}}^{-1} \to v_1^* \omega_{\mathcal{G}_w^{oo}}^{-1}$ which is the transpose of the map $v_1^* \omega_{\mathcal{G}_w^{oo}} \to v_2^* \omega_{\mathcal{G}_w^{oo}}$ obtained by differentiating the dual isogeny: $v_2^*(\mathcal{G}_w^{oo}) \to v_1^* \mathcal{G}_w^{oo}$. The result follows from Lemma 5.4.4. □

Corollary 5.4.5. *Let κ be a weight with $l_w \leq 0$. Let $f \in H^0(X_1^I, \omega^\kappa)$ be such that $Z_w f = \beta_w f$ for some $\beta_w \neq 0$. Then $f = 0$.*

Proof. Since $l_w \leq 0$, we have $Z_w = U_{\mathrm{Kli}(w),1}$ on $H^0(X_1^I, \omega^\kappa)$. Assume that $f \neq 0$, and let n be the order of vanishing of f along $X_1^{=1}$. By considering $\mathrm{Ha}(\mathcal{G}_w)^{-n} f$ and using Proposition 5.4.2 we can suppose that $n = 0$. This contradicts Proposition 5.4.3. □

5.5. **Preliminaries on Kodaira–Spencer.** In this section, we recall the Kodaira–Spencer map and its compatibility with certain functorialities. A convenient reference for what we need is [Lan13].

Let S be a $\mathbf{Z}_{(p)}$-scheme and let X be a smooth S-scheme of relative dimension $3[F : \mathbf{Q}]$. Suppose that we have a tuple (A, ι, λ) with

- A/X an abelian scheme of dimension $2[F : \mathbf{Q}]$.
- $\iota : \mathcal{O}_F \to \mathrm{End}(A) \otimes \mathbf{Z}_{(p)}$ making $\mathrm{Lie}\,(A)$ into a locally free $\mathcal{O}_F \otimes_{\mathbf{Z}} \mathcal{O}_X$-module of rank 2.
- $\lambda : A \to A^t$ a prime to p, $\mathcal{O}_F$-linear quasi-polarization such that $\lambda[p^\infty] : A[p^\infty] \to A^t[p^\infty]$ is an isomorphism.

Then we have the first de Rham cohomology of A/X together with its Hodge filtration

$$0 \to \omega_A \to H^1_{dR}(A/X) \to \omega^\vee_{A^t} \to 0$$

as well as the Gauss–Manin connection

$$H^1_{dR}(A/X) \to H^1_{dR}(A/X) \otimes \Omega^1_{X/S}.$$

Passing to subquotients for the Hodge filtration we obtain the Kodaira–Spencer map for A

$$\omega_A \to \omega^\vee_{A^t} \otimes \Omega^1_{X/S}.$$

The polarization λ induces an isomorphism $\lambda^* : \omega_{A^t} \to \omega_A$. Using this we may obtain a Kodaira–Spencer map for (A, λ)

$$\omega_A \otimes \omega_A \to \Omega^1_{X/S}.$$

Then one checks (see [Lan13, Prop. 6.2.5.18]) that this map factors through the quotient $\mathrm{Sym}^2_{\mathcal{O}_X \otimes \mathcal{O}_F} \omega_A$ of $\omega_A \otimes \omega_A$, so that we obtain a map

$$\mathrm{Sym}^2_{\mathcal{O}_X \otimes \mathcal{O}_F} \omega_A \to \Omega^1_{X/S}.$$

As usual, if Y/S is a smooth scheme of relative dimension d, we write $K_{Y/S}$ for the relative canonical bundle $\wedge^d \Omega^1_{Y/S}$. We will be especially interested in the induced map on top exterior powers

$$\wedge^{3[F:\mathbf{Q}]}(\mathrm{Sym}^2_{\mathcal{O}_X \otimes \mathcal{O}_F} \omega_A) = \det(\omega_A)^3 \to K_{X/S}.$$

Proposition 5.5.1. *Suppose that (A, ι, λ) and (A', ι', λ') are tuples as above and that we have a prime to p quasi-isogeny $\phi : A \to A'$ satisfying $\phi\iota = \iota'\phi$ and $\phi^t \lambda' \phi = x\lambda$ for some $x \in \mathcal{O}_F \otimes \mathbf{Z}_{(p)}$. Then we have a commutative diagram*

$$\begin{tikzcd}
\det(\omega_{A'})^3 \arrow[r] \arrow[d, "\phi^*"] & K_{X/S} \arrow[d, "\cdot N_{F/\mathbf{Q}}(x)^3"] \\
\det(\omega_A)^3 \arrow[r] & K_{X/S}
\end{tikzcd}$$

Proof. It follows from the definitions that under the Kodaira–Spencer maps, $\phi^* : \mathrm{Sym}^2_{\mathcal{O}_X \otimes \mathcal{O}_F} \omega_{A'} \to \mathrm{Sym}^2_{\mathcal{O}_X \otimes \mathcal{O}_F} \omega_A$ induces the endomorphism of $\Omega^1_{X/S}$ given by multiplication by x. The result follows on passing to top exterior powers. □

Proposition 5.5.2. *Let $f : X \to Y$ be a finite flat map of smooth S-schemes of relative dimension $3[F : \mathbf{Q}]$ and let $(A, \iota, \lambda)/Y$ be a tuple as above. Then the Kodaira–Spencer map is compatible with base change in the sense that there is a commutative diagram*

$$\begin{tikzcd}
f^* \det(\omega_A)^3 \arrow[r] \arrow[d, "\wr"'] & f^* K_{Y/S} \arrow[d] \\
\det(\omega_{A_X})^3 \arrow[r] & K_{X/S}
\end{tikzcd}$$

where the horizontal maps are the Kodaira–Spencer maps for A and A_X, the right vertical map is pullback on differentials, and the left vertical map is the natural isomorphism.

Moreover it is compatible with traces in the sense that there is a commutative diagram

$$\begin{tikzcd} f_* \det(\omega_{A_X})^3 \arrow[r] \arrow[d] & f_* K_{X/S} \arrow[d] \\ \det(\omega_A)^3 \arrow[r] & K_{Y/S} \end{tikzcd}$$

where again the horizontal arrows are the Kodaira–Spencer maps for A_X and A while the vertical map on the left comes from the (unnormalized) trace map on functions $f_*\mathcal{O}_X \to \mathcal{O}_Y$ and the isomorphism $\omega_{A_X} \simeq f^*\omega_A$, and the right vertical map is the trace map on dualizing sheaves.

Proof. The commutativity of the first diagram follows from the compatibility of the formation of de Rham cohomology with flat base change, and the compatibility of the Gauss–Manin connection with flat base change (which in turn follows from the compatibility of the Hodge to de Rham spectral sequence with flat base change).

To see that the second diagram commutes, it is by adjunction equivalent to show that the lower square in the following diagram commutes.

$$\begin{tikzcd} f^* \det(\omega_A)^3 \arrow[r] \arrow[d, "\wr"'] & f^* K_{Y/S} \arrow[d] \\ \det(\omega_{A_X})^3 \arrow[r] \arrow[d] & K_{X/S} \arrow[d, "\wr"] \\ f^! \det(\omega_A)^3 \arrow[r] & f^! K_{Y/S} \end{tikzcd}$$

Since we have already seen that the upper square commutes, and since the indicated vertical arrows are isomorphisms, the commutativity of the lower square is equivalent to the commutativity of the outer square. This commutativity follows from unwinding the definitions; indeed, this outer square is the natural one obtained from the Kodaira–Spencer morphisms and the natural transformation from f^* to $f^!$ (which is given by the trace of the morphism f). □

Finally, we recall the Kodaira–Spencer isomorphism for our Shimura varieties.

Proposition 5.5.3. *The Kodaira–Spencer map*

$$\omega^3 \to K_{\mathfrak{Y}^I/\mathbf{Z}_p}$$

is an isomorphism.

Proof. This follows from the usual Kodaira–Spencer isomorphism [Lan13, Thm. 6.4.1.1] and the compatibility with étale base change proved in Proposition 5.5.2 (noting that the formation of the canonical sheaf is compatible with étale base change). □

5.6. **The Hecke operator $U_{\mathrm{Iw}(w),1}$ and traces for partial Frobenius.** We recall the construction of the Hecke operator $U_{\mathrm{Iw}(w),1}$. We have a correspondence (see §5.4, where this correspondence was denoted $\mathfrak{C}^{=_w 2,et}_{\mathrm{Kli}(w),1}$ but we adopt here a

simplified notation $\mathfrak{C}_w^I$)

$$\begin{tikzcd} & \mathfrak{C}_w^I \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ \mathfrak{Y}^{I,=_w 2} & & \mathfrak{Y}^{I,=_w 2} \end{tikzcd}$$

where $\mathfrak{C}_w^I$ parameterizes a point $(A, \iota, \lambda, \{H_v\}_{v \in S_p}, \eta)$ of $\mathfrak{Y}^{I,=_w 2}$ along with an étale maximal isotropic subgroup $L_w \subset A[w]$. The map p_1 simply forgets L_w. To describe p_2, consider the étale isogeny $\pi : A \to A/L_w$. Then p_2 sends $(A, \iota, \lambda, \{H_v\}_{v \in S_p})$ to A/L_w with the induced action of $\mathcal{O}_F \otimes \mathbf{Z}_{(p)}$, the prime to p quasi-polarization obtained by descending $x_w \lambda$, and the level structures $\pi(H_v)$ and $\pi(\eta)$. Since the subgroup $A[x_w]/L_w$ of A/L_w is the canonical multiplicative subgroup of $A[x_w]$, we see that p_2 is an isomorphism.

For any weight κ for G, pullback by the universal étale isogeny over $\mathfrak{C}_w^I$ induces an isomorphism of sheaves $p_2^* \omega^\kappa \to p_1^* \omega^\kappa$, and the Hecke operator $U_{\mathrm{Iw}(w),1}$ is obtained from the composition of maps of sheaves over $\mathfrak{Y}^{I,=_w 2}$

$$p_{1,*} p_2^* \omega^\kappa \to p_{1,*} p_1^* \omega^\kappa \overset{\text{“}\frac{1}{p^3}\mathrm{Tr}_{p_1}\text{”}}{\to} \omega^\kappa.$$

Now we turn to the Kodaira–Spencer isomorphism $\omega^3 \simeq K_{\mathfrak{Y}^{I,=_w 2}}$.

Proposition 5.6.1. *There is a commutative diagram of sheaves on* $\mathfrak{Y}^{I,=_w 2}$

$$\begin{tikzcd} p_{1,*} p_2^* \omega^3 \arrow[r] \arrow[d] & \omega^3 \arrow[d] \\ p_{1,*} p_2^* K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p} \arrow[r, "\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3} tr"] & K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p} \end{tikzcd}$$

where the vertical arrows are the Kodaira–Spencer isomorphism, and the top horizontal arrow is $U_{\mathrm{Iw}(w),1}$. *The bottom horizontal arrow is defined as follows: since* p_2 *is an isomorphism, we may identify* $p_2^* K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p}$ *with* $K_{\mathfrak{C}_w^I/\mathbf{Z}_p}$, *and the morphism then comes from the trace map for* p_1 *on dualizing sheaves, multiplied by a factor of* $\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3} \in \mathbf{Z}_{(p)}^\times$.

Proof. This follows from Propositions 5.5.1 and 5.5.2. □

We note that although we are primarily interested in using Proposition 5.6.1 on the special fibre, we cannot apply Propositions 5.5.1 and 5.5.2 directly on the special fibre because some of the maps in the commutative square reduce to 0 modulo p.

We may also describe $U_{\mathrm{Iw}(w),1}$ in weights other than parallel weight 3 using traces on differentials. For any weight $\kappa = (k_v, l_v)_{v \in S_p}$ we let $\kappa - 3 = (k_v - 3, l_v - 3)$. Then tensoring the Kodaira–Spencer isomorphism with $\omega^{\kappa-3}$ we have an isomorphism $\omega^\kappa \simeq K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p} \otimes \omega^{\kappa-3}$. Then we have a commutative diagram of sheaves on $\mathfrak{Y}^{I,=_w 2}$

$$\begin{tikzcd} p_{1,*} p_2^* \omega^\kappa \arrow[rr] \arrow[d] & & \omega^\kappa \arrow[d] \\ p_{1,*} p_2^* (K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p} \otimes \omega^{\kappa-3}) \arrow[r] & (p_{1,*} p_2^* K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p}) \otimes \omega^{\kappa-3} \arrow[r] & K_{\mathfrak{Y}^{I,=_w 2}/\mathbf{Z}_p} \otimes \omega^{\kappa-3} \end{tikzcd}$$

where on the bottom row, the first map is an isomorphism coming from the projection formula and the isomorphism $p_2^*\omega^{\kappa-3} \simeq p_1^*\omega^{\kappa-3}$ and the second map is the tensor product of the map of Proposition 5.6.1 and the identity.

We would now like to understand the behavior of $U_{\mathrm{Iw}(w),1}$ beyond the w-ordinary locus on the special fibre. In order to do this we make the following definition.

Definition 5.6.2. We define a "partial Frobenius" map

$$F_w : Y_1^I \to Y_1^I$$

as follows: given a point $(A, \iota, \lambda, \{H_v\}_{v\in S_p}, \eta)$ of Y_1^I, we may consider the maximal isotropic subgroup $L_w \subset A[w]$ defined by

$$L_w = \ker(F : A[w^\infty] \to A[w^\infty]^{(p)})$$

and form the subgroup of degree $p^{4[F:\mathbf{Q}]-2}$

$$\tilde{L}_w = L_w \times \prod_{v\neq w} A[v] \subset A$$

and the isogeny $\pi : A \to \tilde{A} = A/\tilde{L}_w$. $\tilde{L}_w$ is isotropic for the polarization $\frac{p^2}{x_w}\lambda$ which thus descends to a principal polarization $\tilde{\lambda}$ on $A/\tilde{L}_w$. Then the map F_w is defined by

$$F_w(A, \iota, \lambda, \{H_v\}_{v\in S_p}, \eta) = (\tilde{A}, \tilde{\iota}, \tilde{\lambda}, \{\tilde{H}_v\}_{v\in S_p}, \tilde{\eta})$$

where $\tilde{A}$, and $\tilde{\lambda}$ are as described above, $\tilde{\iota}$ is the induced action of $\mathcal{O}_F \otimes \mathbf{Z}_{(p)}$, $\tilde{H}_v = \pi H_v$ for $v \neq w$, $\tilde{H}_w = H_w^{(p)} \subset A[w^\infty]^{(p)} \simeq (A/L_w)[w^\infty]$, and $\tilde{\eta} = \frac{1}{p}\pi\eta$.

Note that this definition depends on the choice of x_w.

To explain why we call the map F_w a partial Frobenius, observe that according to the product decomposition

$$A[p^\infty] = \prod_{v|p} A[v^\infty]$$

we have $\tilde{A}[w^\infty] = \tilde{A}[w^\infty]/L_w \simeq A[w^\infty]^{(p)}$ while $\tilde{A}[v^\infty] \simeq A[v^\infty]$ for all $v \neq w$. In particular, according to the local product structure of Y_1^I coming from the Serre–Tate theorem and the product decomposition of the p-divisible group $A[p^\infty]$, F_w looks like Frobenius on the factor corresponding to w.

As a consequence of this we may record

Proposition 5.6.3. *F_w is finite flat of degree p^3. It restricts to a map $F_w : Y_1^{I,=_w 1} \to Y_1^{I,=_w 1}$ which is finite flat of degree p^2.*

Proof. Using the Serre–Tate theorem and the description of F_w on the p-divisible group above, this follows from the fact that Frobenius on a smooth variety of dimension n is finite flat of degree p^n. □

The identification $\tilde{A}[w^\infty] \simeq A[w^\infty]^{(p)}$ induces a canonical isomorphism $F_w^*\omega_w \simeq \omega_w^p$ while the isomorphisms $\tilde{A}[v^\infty] \simeq A[v^\infty]$ for $v \neq w$ induce canonical isomorphisms $F_w^*\omega_v \simeq \omega_v$.

The point of this definition is that if we identify $\mathfrak{C}_w^I$ with $\mathfrak{Y}^{I,=_w 2}$ via p_2, the map p_1 on the special fibre is simply the partial Frobenius F_w restricted to $Y_1^{I,=_w 2}$. Moreover making these identifications, the isogeny $p_1^*\mathcal{G}_w \to p_2^*\mathcal{G}_w$ becomes V :

$\mathcal{G}_w^{(p)} \to \mathcal{G}_w$ (as its dual is Frobenius) and so the pullback map $p_2^*\omega_w \to p_1^*\omega_w$ becomes $\mathrm{Ha}_w : \omega_w \to \omega_w^p$.

As in §3.8.16, we may consider trace maps for F_w on differentials $F_{w,*}K_{Y_1^I} \to K_{Y_1^I}$. Tensoring with any line bundle $\mathcal{L}$ on Y_1^I and using the projection formula $F_{w,*}K_{Y_1^I} \otimes \mathcal{L} \simeq F_{w,*}(K_{Y_1^I} \otimes F_w^*\mathcal{L})$, we obtain a twisted trace map

$$F_{w,*}(K_{Y_1^I} \otimes F_w^*\mathcal{L}) \to K_{Y_1^I} \otimes \mathcal{L}.$$

We may similarly consider twisted trace maps for line bundles on the divisor $Y_1^{I,=_w 1}$.

Now we restrict to parallel weight 2 and work on the special fibre. With the identifications we have made, our discussion above shows that we have a commutative diagram of sheaves on $Y_1^{I,=_2 w}$

$$\begin{tikzcd}
F_{w,*}\omega^2 \arrow[rr, "U_{\mathrm{Iw}(w),1}"] \arrow[d] & & \omega^2 \arrow[d] \\
F_{w,*}(K_{Y_1^{I,=_w 2}} \otimes \omega^{-1}) \arrow[r, "\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3}\mathrm{Ha}_w^{-1}"] & F_{w,*}(K_{Y_1^{I,=_w 2}} \otimes \omega^{-1} \otimes \omega_w^{1-p}) \arrow[r] & K_{Y_1^{I,=_w 2}} \otimes \omega^{-1}
\end{tikzcd}$$

Now we want to extend this description to all of Y_1^I. Here is the first main result of this section.

Proposition 5.6.4. *The map* $\mathrm{Ha}_w \cdot U_{\mathrm{Iw}(w),1} : F_{w,*}\omega^2 \to \omega^2 \otimes \omega_w^{p-1}$ *of sheaves on* $Y_1^{I,=_w 2}$ *extends to* Y_1^I *and fits in to a commutative diagram of sheaves on* Y_1^I

$$\begin{tikzcd}
F_{w,*}\omega^2 \arrow[rr, "\mathrm{Ha}_w \cdot U_{\mathrm{Iw}(w),1}"] \arrow[d] & & \omega^2 \otimes \omega_w^{p-1} \arrow[d] \\
F_{w,*}(K_{Y_1^{I,=_w 2}} \otimes \omega^{-1}) \arrow[r, "\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3}\mathrm{Ha}_w^{p-1}"] & F_{w,*}(K_{Y_1^{I,=_w 2}} \otimes \omega^{-1} \otimes \omega_w^{(p-1)^2}) \arrow[r] & K_{Y_1^{I,=_w 2}} \otimes \omega_w^{p-1} \otimes \omega^{-1}
\end{tikzcd}$$

Proof. To prove the proposition, it suffices to establish the commutativity of the diagram over $Y_1^{I,=_w 2}$, as the vertical maps are isomorphisms, and the maps on the bottom are already defined over Y_1^I. This commutativity follows from the discussion above and the fact that $F_w^*\mathrm{Ha}_w = \mathrm{Ha}_w^p$. □

Now we are going to restrict to the divisor $Y_1^{I,=_w 1}$. The Kodaira–Spencer isomorphism $K_{Y_1^I} \simeq \omega^3$ of Proposition 5.5.3 induces by the adjunction formula an isomorphism

$$K_{Y_1^{I,=_w 1}} \simeq \omega^3 \otimes \omega_w^{p-1}|_{Y_1^{I,=_w 1}}.$$

Proposition 5.6.5. *There is a commutative diagram of sheaves on* Y_1^I

$$\begin{tikzcd}
F_{w,*}(K_{Y_1^I} \otimes \omega^{-1}) \arrow[r, "\mathrm{Ha}_w^{p-1}"] \arrow[d] & F_{w,*}(K_{Y_1^I} \otimes \omega^{-1} \otimes \omega_w^{(p-1)^2}) \arrow[r] & K_{Y_1^I} \otimes \omega^{-1} \otimes \omega_w^{p-1} \arrow[d] \\
F_{w,*}(K_{Y_1^{I,=_w 1}} \otimes \omega^{-1} \otimes \omega_w^{1-p}) \arrow[rr] & & K_{Y_1^{I,=_w 1}} \otimes \omega^{-1}
\end{tikzcd}$$

where the vertical maps are obtained from restriction and the adjunction formula as recalled above, and the bottom horizontal map is a twisted trace for F_w *on the divisor* $Y_1^{I,=_w 1}$ *and the line bundle* ω^{-1}.

Proof. The commutativity of this diagram follows from Proposition 3.8.17, where in the notation of that proposition we take $X = Y = Y_1^I$, $D' = D = Y_1^{I,=_w 1}$, $f = F_w$ and $n = p$, and identify $\mathcal{O}_{Y_1^I}(Y_1^{I,=_w 1})$ with ω_w^{p-1} via Ha_w (tensor the commutative diagram of Proposition 3.8.17 with $\omega^{-1} \otimes \omega_w^{p-1}$). $\square$

We may then define a map $\gamma_1 : F_{w,*}(\omega^2|_{Y_1^{I,=_w 1}}) \to \omega^2 \otimes \omega_w^{p-1}|_{Y_1^{I,=_w 1}}$ by the diagram

$$\begin{tikzcd} F_{w,*}(\omega^2|_{Y_1^{I,=_w 1}}) \arrow[r, "\gamma_1"] \arrow[d, "\wr"'] & \omega^2 \otimes \omega_w^{p-1}|_{Y_1^{I,=_w 1}} \arrow[d, "\wr"] \\ F_{w,*}(K_{Y_1^{I,=_w 1}} \otimes \omega^{-1} \otimes \omega_w^{1-p}) \arrow[r] & K_{Y_1^{I,=_w 1}} \otimes \omega^{-1} \end{tikzcd} \tag{5.6.6}$$

where the vertical maps are Kodaira–Spencer and the bottom horizontal map is $\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3}$ times the twisted trace for ω^{-1}.

Now combining Proposition 5.6.4 with Proposition 5.6.5 with the definition of γ_1 by (5.6.6) we have proved the following.

Proposition 5.6.7. *There is a commutative diagram*

$$\begin{tikzcd} H^0(Y_1^I, \omega^2) \arrow[r, "\mathrm{Ha}_w \cdot U_{\mathrm{Iw}(w),1}"] \arrow[d] & H^0(Y_1^I, \omega^2 \otimes \omega_w^{p-1}) \arrow[d] \\ H^0(Y_1^{I,=_w 1}, \omega^2|_{Y_1^{I,=_w 1}}) \arrow[r, "\gamma_1"] & H^0(Y_1^{I,=_w 1}, \omega^2 \otimes \omega_w^{p-1}|_{Y_1^{I,=_w 1}}) \end{tikzcd}$$

where the vertical maps are restrictions and the horizontal maps are as explained above.

5.7. **The Hecke operator $U_{w,2}$ on the w-non ordinary locus.** In this section we consider the Hecke operator $U_{w,2}$ that was first introduced in §4.5.8. We consider the correspondence

$$\begin{tikzcd} & \mathfrak{C}_{w,2}^I \arrow[dl, "p_2"'] \arrow[dr, "p_1"] & \\ \mathfrak{Y}^I & & \mathfrak{Y}^I \end{tikzcd}$$

which is the composition of the correspondences $\mathfrak{C}_{w,2,1}(p)$ and $\mathfrak{C}_{w,2,2}(p)$ considered in §4.5.8 (or more precisely their restrictions to the interior of the moduli space). The correspondence $\mathfrak{C}_{w,2}$ admits the following direct description: it parametrizes isogenies $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ whose kernel K_w is a totally isotropic subgroup of $\mathcal{G}_w[p^2]$ which has trivial intersection with the group $p_1^*H_w$. To see this, note that K_w fits into an exact sequence $0 \to K_w[p] \to K_w \to K_w/K_w[p] \to 0$ where $K_w[p]$ is a finite flat group scheme of rank p^3 and étale rank p, and $K_w/K_w[p]$ is a finite étale group scheme of rank p.

There is yet another description of $\mathfrak{C}_{w,2}^I$ that will be important for us. To any point $(\mathcal{G}, \iota, \lambda, \{H_v\}_{v \in S_p}, \eta) \in \mathfrak{Y}^I$ we can associate a subgroup $\tilde{L}_w \subset \mathcal{G}[p^2]$ as follows: the finite flat group scheme $x_w^{-1}H_w/H_w^\perp \subset \mathcal{G}_w/H_w^\perp$ of degree p^2 contains a canonical multiplicative subgroup L'_w of degree p (as $x^{-1}H_w/\mathcal{G}_w[p] \simeq H_w$ is multiplicative and $\mathcal{G}_w[p]/H_w^\perp \simeq H_w^D$ is étale, we see that $x_w^{-1}H_w/H_w^\perp$ is isomorphic at geometric

points to $\mu_p \times \mathbf{Z}/p\mathbf{Z}$, and hence over the entire (reduced) special fibre, the kernel of Frobenius on $x_w^{-1}H_w/H_w^{\perp}$ is a multiplicative group of order p which lifts uniquely over $\mathfrak{Y}^I$). Then we may define the group L_w to be the preimage of L'_w under the isogeny $A \to A/H_w^{\perp}$. Observe that $L_w \subset A[w^2]$ is a totally isotropic subgroup of degree p^4. Then we take

$$\tilde{L}_w = L_w \times \prod_{v \neq w} A[v^2].$$

We temporarily write $\mathfrak{Y}^I_{w-\mathrm{sph}}$ for the formal completion of $Y_{K_pK^p}$ with $K_p = \mathrm{GSp}_4(\mathcal{O}_{F_w}) \prod_{v \in I, v \neq w} \mathrm{Kli}(v) \prod_{v \in I^c} \mathrm{Iw}(v)$, along the open subvariety of the special fibre where $A[w^\infty]$ has p-rank ≥ 1, H_v is multiplicative for $v \in I$, $v \neq w$, and L_v is multiplicative for $v \in I^c$. Then there is a natural map $f : \mathfrak{Y}^I \to \mathfrak{Y}^I_{w-\mathrm{sph}}$ which forgets the Klingen level structure H_w at w. It is étale and affine (see Proposition 4.3.3).

We define a map $\psi_w : \mathfrak{Y}^I \to \mathfrak{Y}^I_{w-\mathrm{sph}}$ by sending a point $(\mathcal{G}, \iota, \lambda, \{H_v\}_{v \in S_p}, \eta)$ to $\mathcal{G}/\tilde{L}_w$ with the polarization descended from $\frac{p^4}{x_w^2}\lambda$, the induced action of $\mathcal{O}_F$ and the level structures $p^{-2}\pi\eta$ and πH_v for $v \neq w$ where $\pi : \mathcal{G} \to \mathcal{G}/\tilde{L}_w$ is the isogeny.

Lemma 5.7.1. *The correspondence $\mathfrak{C}^I_{w,2}$ fits in the following Cartesian diagram*

$$\begin{tikzcd} \mathfrak{C}^I_{w,2} \arrow[r, "p_1"] \arrow[d, "p_2"] & \mathfrak{Y}^I \arrow[d, "f"] \\ \mathfrak{Y}^I \arrow[r, "\psi_w"] & \mathfrak{Y}^I_{w-\mathrm{sph}} \end{tikzcd}$$

Proof. Let $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ be the universal isogeny over $\mathfrak{C}^I_{w,2}$. Then the composite of this isogeny with the isogeny $p_2^*\mathcal{G} \to p_2^*\mathcal{G}/\tilde{L}_w$ identifies with multiplication by p^2 on $p_1^*\mathcal{G}$. □

Lemma 5.7.2. *The map $p_1 : \mathfrak{C}^I_{w,2} \to \mathfrak{Y}^I$ is finite flat of degree p^4.*

Proof. The correspondence $\mathfrak{C}^I_{w,2}$ is smooth, and the map p_1 is generically étale of degree p^4 and finite. It follows from miracle flatness that it is finite flat. □

We can now deduce the following important relation between $\mathfrak{C}^I_{w,2}$ and F_w^2 over $Y_1^{I,=_w 1}$.

Proposition 5.7.3. *The restriction of p_2 to the scheme theoretic preimage $p_2^{-1}(Y_1^{I,=_w 1})$ is an isomorphism to $Y_1^{I,=_w 1}$. Making this identification, p_1 becomes $F_w^2 : Y_1^{I,=_w 1} \to Y_1^{I,=_w 1}$.*

Proof. We observe that the restriction of f to $Y_1^{I,=_w 1}$ is an isomorphism. This implies that the restriction of p_2 is an isomorphism. Now we examine the definition of $\tilde{L}_w$. The key observation is that L_w coincides with the kernel of $\mathrm{Frob}^2 : \mathcal{G}_w \to \mathcal{G}_w^{(p^2)}$. It suffices to check this on geometric points. Over a geometric point of $Y_1^{I,=_w 1}$, the p-divisible group $\mathcal{G}_w$ has a decomposition into a product of multiplicative, slope $\frac{1}{2}$ and étale group p-divisible groups: $\mathcal{G}_w = \mathcal{G}_w^{et} \times \mathcal{G}_w^{oo} \times \mathcal{G}_w^{m}$. Then the kernel of Frob^2 is simply $\mathcal{G}_w^{oo}[p] \times \mathcal{G}_w^{m}[p^2]$, and this group equals L_w. □

Pullback by the universal isogeny over $\mathfrak{C}^I_{w,2}$ induces a morphism $\delta_0 : p_2^*\omega_w \to p_1^*\omega_w$, as well as isomorphisms $p_2^*\omega_v \to p_1^*\omega_v$ for $v \neq w$. The following proposition is implicitly contained in Lemma 4.6.4, but we briefly recall the argument.

Proposition 5.7.4.

(1) The map δ_0 is divisible by p and the resulting map $\delta = \frac{1}{p}\delta_0 : p_2^*\omega_w \to p_1^*\omega_w$ is an isomorphism.
(2) Under the isomorphism $\delta^{p-1} : p_2^*\omega_w^{p-1} \simeq p_1^*\omega_w^{p-1}$ we have $p_1^*\mathrm{Ha}_w = p_2^*\mathrm{Ha}_w$.

Proof. Because $\mathfrak{C}^I_{w,2}$ is smooth we are free to check the first claim on the ordinary locus where it simply follows from the fact that the isogeny $p_1^*\mathcal{G} \to p_2^*\mathcal{G}$ has kernel of multiplicative rank one. The second claim follows from [Pil20, Lem. 10.5.2.1]. □

Making the identifications of the Proposition 5.7.3, we may view the restriction of δ to $Y_1^{I,=_w 1}$ as an isomorphism

$$\delta|^{-1}_{Y_1^{I,=_w 1}} : \omega_w \to (F_w^2)^*\omega_w \simeq \omega_w^{p^2}$$

or equivalently as a non vanishing section $\delta|^{-1}_{Y_1^{I,=_w 1}} \in H^0(Y_1^{I,=_w 1}, \omega_w^{p^2-1})$. We also denote by $\delta' : \prod_{v\neq w} p_2^*\omega_w \simeq \prod_{v\neq w} p_1^*\omega_w$ the isomorphism coming from the pullback of differentials.

In weights $\kappa = (k_v, l_v)$ with $l_w \geq 0$, the Hecke operator

$$U_{w,2} : p_{1,*}p_2^*\omega^\kappa \to \omega^\kappa$$

is defined by tensoring the unnormalized trace map $p_{1,*}p_2^*\mathcal{O}_{Y_1^I} \to \mathcal{O}_{Y_1^I}$ with the unnormalized pullback map $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$, and normalizing by a factor of $\frac{1}{p^{3+l_w}}$ (see §4.5.8; equivalently, the normalized map $p_2^*\omega^\kappa \to p_1^*\omega^\kappa$ is constructed with the help of the operator δ).

First we may use the Kodaira–Spencer isomorphism to describe $U_{w,2}$ in weight 3 in terms of traces on differentials.

Proposition 5.7.5. *There is a commutative diagram of sheaves on $\mathfrak{Y}^I$*

$$\begin{tikzcd} p_{1,*}p_2^*\omega^3 \arrow[r, "U_{w,2}"] \arrow[d] & \omega^3 \arrow[d] \\ p_{1,*}p_2^*K_{\mathfrak{Y}^I/\mathbf{Z}_p} \arrow[r, "\frac{N_{F/\mathbf{Q}}(x_w)^6}{p^6}\mathrm{tr}"] & K_{\mathfrak{Y}^I/\mathbf{Z}_p} \end{tikzcd}$$

where the vertical arrows are the Kodaira–Spencer isomorphism, and the bottom horizontal arrow is defined as follows: since p_2 is étale, we may identify $p_2^*K_{\mathfrak{Y}^I/\mathbf{Z}_p}$ with $K_{\mathfrak{C}^I_{w,2}/\mathbf{Z}_p}$, and the morphism then comes from the trace map for p_1 on dualizing sheaves, multiplied by a factor of $\frac{N_{F/\mathbf{Q}}(x_w)^6}{p^6} \in \mathbf{Z}^\times_{(p)}$.

Proof. This follows from Propositions 5.5.1 and 5.5.2. □

In parallel weight 2 we can still express the cohomological correspondence $U_{w,2}$ by using a similar commutative diagram of sheaves on $\mathfrak{Y}^I$:

$$\begin{tikzcd}
p_{1,*}p_2^*\omega^2 \arrow[r, "U_{w,2}"] \arrow[d] & \omega^2 \arrow[d] \\
p_{1,*}p_2^*(K_{\mathfrak{Y}^I/\mathbf{Z}_p}\otimes\omega^{-1}) \arrow[r, "\frac{N_{F/\mathbf{Q}}(x_w)^6}{p^6}\mathrm{tr}\otimes(\delta'\delta)^{-1}"] & K_{\mathfrak{Y}^I/\mathbf{Z}_p}\otimes\omega^{-1}
\end{tikzcd}$$

We can restrict to $Y^{I,=_w 1}$ and obtain the following:

Proposition 5.7.6. *There is a commutative diagram*

$$\begin{tikzcd}
p_{1,*}p_2^*(K_{Y_1^I}\otimes\omega^{-1}) \arrow[r] \arrow[d] & K_{Y_1^I}\otimes\omega^{-1} \arrow[d] \\
p_{1,*}p_2^*(K_{Y_1^{I,=_w 1}}\otimes\mathcal{O}_{Y_1^I}(-Y_1^{I,=_w 1})|_{Y_1^{I,=_w 1}}\otimes\omega^{-1}) \arrow[r] \arrow[d] & K_{Y_1^{I,=_w 1}}\otimes\mathcal{O}_{Y_1^I}(-Y_1^{I,=_w 1})|_{Y_1^{I,=_w 1}}\otimes\omega^{-1} \arrow[d] \\
(F_w^2)_*(K_{Y_1^{I,=_w 1}}\otimes\omega^{-1}\otimes\omega_w^{1-p}) \arrow[r] & K_{Y_1^{I,=_w 1}}\otimes\omega^{-1}\otimes\omega_w^{1-p}
\end{tikzcd}$$

where:

- *The upper vertical maps are obtained by restriction to $Y_1^{I,=_w 1}$ and the adjunction isomorphism $K_{Y_1^I}|_{Y^{I,=_w 1}}\simeq K_{Y_1^{I,=_w 1}}\otimes\mathcal{O}_{Y_1^I}$.*
- *The lower vertical maps are obtained from making the identification of p_2 with* id *and p_1 with F_w^2 of Proposition 5.7.3 as well as using the isomorphism $\mathrm{Ha}_w:\mathcal{O}(Y_1^{I,=_w 1})\to\omega_w^{p-1}$.*
- *The top horizontal arrow is the composition of $(\delta\delta')^{-1}$ and the twisted trace for ω^{-1} on $K_{Y_1}^I$, as on the bottom row of the diagram immediately preceding this proposition.*
- *The middle horizontal arrow is multiplication by $(\delta\delta')|_{Y_1^{I,=_w 1}}^{-1}$ followed by the twisted trace for p_1 on the sheaf $\mathcal{O}_{Y_1^I}(-Y_1^{I,=_w 1})\otimes\omega^{-1}$.*
- *The bottom horizontal arrow is multiplication by $\delta|_{Y_1^{I,=_w 1}}^{-p}\delta'|_{Y_1^{I,=_w 1}}^{-1}$ followed by the twisted trace for F_w^2 on the line bundle $\omega^{-1}\otimes\omega_w^{1-p}$ (normalized by the p-adic unit $\frac{N_{F/\mathbf{Q}}(x_w)^6}{p^6}$).*

Proof. The commutativity of the top square follows from Proposition 3.8.17 while the commutativity of the bottom square follows from Proposition 5.7.4. The reason we get multiplication by $\delta|_{Y_1^{I,=_w 1}}^{-p}\delta'|_{Y_1^{I,=_w 1}}^{-1}$ in the bottom horizontal arrow is because we multiply the original $(\delta\delta')|_{Y_1^{I,=_w 1}}^{-1}$ with $\delta|_{Y_1^{I,=_w 1}}^{-1-p}$ which arises when relating the isomorphisms $p_i^*(\mathrm{Ha}_w:\mathcal{O}(Y_1^{I,=_w 1})\to\omega_w^{p-1})$ for $i=1,2$. $\square$

Our goal from now on is to interpret the bottom horizontal line of this diagram in terms of the map γ_1 of (5.6.6). We introduce a map $\gamma_2: F_{w,*}(\omega^2\otimes\omega_w^{p-1}|_{Y_1^{I,=_w 1}})\to$

$\omega^2|_{Y_1^{I,=w1}}$ defined by the diagram

(5.7.7)

$$\begin{tikzcd}
F_{w,*}(\omega^2 \otimes \omega_w^{p-1}|_{Y_1^{I,=w1}}) \arrow[rr, "\gamma_2"] \arrow[d, "\wr"'] & & \omega^2|_{Y_1^{I,=w1}} \arrow[d, "\wr"] \\
F_{w,*}(K_{Y_1^{I,=w1}} \otimes \omega^{-1}) \arrow[rr, bend right, "\delta|_{Y_1^{I,=w1}}^{-1}"'] & F_{w,*}(K_{Y_1^{I,=w1}} \otimes \omega^{-1} \otimes \omega_w^{1-p^2}) \arrow[r] & K_{Y_1^{I,=w1}} \otimes \omega^{-1} \otimes \omega_w^{1-p}
\end{tikzcd}$$

where the vertical maps are induced by Kodaira–Spencer, and on the bottom the first horizontal map is multiplication by $\delta|_{Y_1^{I,=w1}}^{-1}$, while the second is the twisted trace for $\mathcal{L} = \omega^{-1} \otimes \omega_w^{1-p}$ (normalized by the p-adic unit $\frac{N_{F/\mathbf{Q}}(x_w)^3}{p^3}$).

We now consider the composition $\gamma_2 \circ \gamma_1 : (F_w^2)_*(\omega^2|_{Y_1^{I,=w1}}) \to \omega^2|_{Y_1^{I,=w1}}$.

Proposition 5.7.8. *There is a commutative diagram*

$$\begin{tikzcd}
(F_w^2)_*(\omega^2|_{Y_1^{I,=w1}}) \arrow[r, "\gamma_1"] \arrow[d, "\wr"'] & F_{w,*}(\omega^2 \otimes \omega_w^{p-1}|_{Y_1^{I,=w1}}) \arrow[r, "\gamma_2"] & \omega^2|_{Y_1^{I,=w1}} \arrow[d, "\wr"] \\
(F_w^2)_*(K_{Y_1^{I,=w1}} \otimes \omega^{-1} \otimes \omega_w^{1-p}) \arrow[rr, bend right, "\delta^{-p}\delta'^{-1}|_{Y_1^{I,=w1}}"'] & (F_w^2)_*(K_{Y_1^{I,=w1}} \otimes \omega^{-1} \otimes \omega_w^{1-p^3}) \arrow[r] & K_{Y_1^{I,=w1}} \otimes \omega^{-1} \otimes \omega_w^{1-p}
\end{tikzcd}$$

where the vertical arrows are given by the Kodaira–Spencer isomorphism and on the bottom row we first multiply by $\delta|_{Y_1^{I,=w1}}^{-p}$ and then take a twisted trace for F_w^2 (normalized by the p-adic unit $\frac{N_{F/\mathbf{Q}}(x_w)^6}{p^6}$) and the line bundle $\mathcal{L} = \omega^{-1} \otimes \omega_w^{1-p}|_{Y_1^{I,=w1}}$.

Proof. This follows from the fact that $F_w^*\delta|_{Y_1^{I,=w1}} = \delta|_{Y_1^{I,=w1}}^p$. □

Combining Proposition 5.7.6 with Proposition 5.7.8 we have proved the following:

Proposition 5.7.9. *There is a commutative diagram*

$$\begin{tikzcd}
H^0(Y_1^I, \omega^2) \arrow[r, "U_{w,2}"] \arrow[d] & H^0(Y_1^I, \omega^2) \arrow[d] \\
H^0(Y_1^{I,=w1}, \omega^2|_{Y_1^{I,=w1}}) \arrow[r, "\gamma_2 \circ \gamma_1"] & H^0(Y_1^{I,=w1}, \omega^2|_{Y_1^{I,=w1}})
\end{tikzcd}$$

where the vertical maps are restrictions and the horizontal maps are as explained above.

5.8. **Main doubling results.** There is an obvious injective restriction map:

$$H^0(X_1^I, \omega^2(-D)) \to H^0(X_1^{I,=w2}, \omega^2(-D))$$

which is equivariant for the action of the Hecke algebra away from w, and for the actions of $U_{w,2}$ and $U_{w,0}$. We now compare the action of $U_{\mathrm{Kli}(w),1}$, which acts on both the left hand and right hand modules, and $U_{\mathrm{Iw}(w),1}$, which acts on the right hand module.

We have defined a Hecke operator Z_w on $H^0(X_1^{I,=_w 2}, \omega^2(-D))$ with $U_{\mathrm{Kli}(w),1} = U_{\mathrm{Iw}(w),1} + Z_w$ (see §5.4).

Lemma 5.8.1. *On $H^0(X_1^{I,=_w 2}, \omega^2(-D))$ we have the identity of operators $U_{\mathrm{Iw}(w),1} Z_w = U_{w,2}$.*

Proof. This is immediate from Lemma 4.5.17.

□

We introduce the doubling map:

$$\begin{array}{rcl} H^0(X_1^I, \omega^2(-D)) \oplus H^0(X_1^I, \omega^2(-D)) & \to & H^0(X_1^{I,=_w 2}, \omega^2(-D)) \\ (f, g) & \mapsto & f + Z_w g \end{array}$$

In this formula f and g on the right hand side are viewed as sections of $H^0(X_1^{I,=_w 2}, \omega^2(-D))$ via the above restriction map.

We can define an operator that we formally denote by $U_{\mathrm{Iw}(w),1}$ on the left hand side by the following matrix:

$$U_{\mathrm{Iw}(w),1} = \begin{pmatrix} U_{\mathrm{Kli}(w),1} & U_{w,2} \\ -1 & 0 \end{pmatrix}$$

Lemma 5.8.2. *The doubling map is equivariant for the action of $U_{\mathrm{Iw}(w),1}$. The operator $U_{\mathrm{Iw}(w),1}$ on $H^0(X_1^I, \omega^2(-D)) \oplus H^0(X_1^I, \omega^2(-D))$ commutes with the action of $U_{w,2}$.*

Proof. The equivariance follows from Lemma 5.8.1, and the commutativity follows from Lemma 4.5.15. □

We now consider the $U_{w,2}$-ordinary part:

$$e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)) \oplus e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)).$$

We have restricted to the direct factor $H^0(X_1^{G_1,I}, \omega^2(-D))$ of $H^0(X_1^I, \omega^2(-D))$ in order to be able to use local finiteness and apply ordinary projectors.

Lemma 5.8.3.

(1) The image of

$$e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)) \oplus e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D))$$

via the doubling map lands in $e(U_{\mathrm{Iw}(w),1} U_{w,2}) H^0(X_1^{G_1,I,=_w 2}, \omega^2(-D))$.

(2) The operator $U_{\mathrm{Iw}(w),1}$ is (left and right) invertible on

$$e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)) \oplus e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)).$$

(3) On $e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D)) \oplus e(U_{w,2}) H^0(X_1^{G_1,I}, \omega^2(-D))$ we have the identity $U_{\mathrm{Kli}(w),1} = U_{\mathrm{Iw}(v),1} + U_{w,2} U_{\mathrm{Iw}(v),1}^{-1}$. In particular $U_{\mathrm{Kli}(w),1}$ and $U_{\mathrm{Iw}(w),1}$ commute with each other.

Proof. The doubling map

$$H^0(X_1^{G_1,I}, \omega^2(-D)) \oplus H^0(X_1^{G_1,I}, \omega^2(-D)) \to H^0(X_1^{G_1,I,=_w 2}, \omega^2(-D))$$

can be written as an inductive limit of maps between finite dimensional vector spaces stable under the $U_{\mathrm{Iw}(w),1}$ and $U_{w,2}$ operators, so we will freely use the usual properties of linear endomorphisms on finite dimensional vector spaces. We first

observe that the operator $U_{w,2}$ is invertible on $e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D))$. Concretely, for any $f \in e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D))$ we have $e(U_{w,2})f = f = U_{w,2}^{p^{N!}}f$ for N large enough, and $U_{w,2}^{-1}f = U_{w,2}^{p^{N!}-1}f$. Therefore we may consider the operator

$$U_{w,2}^{-1}\begin{pmatrix} 0 & -U_{w,2} \\ 1 & U_{\mathrm{Kli}(w),1} \end{pmatrix}$$

on $e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D)) \oplus e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D))$, and it is straightforward to check (using that $U_{\mathrm{Kli}(w),1}$ and $U_{w,2}$ commute, as we noted in the proof of Lemma 5.8.2) that this is a 2-sided inverse of $U_{\mathrm{Iw}(w),1}$. This proves the first and second points. The third point is obvious from the formulae defining $U_{\mathrm{Iw}(w),1}$ and $U_{\mathrm{Iw}(w),1}^{-1}$. □

We now prove our doubling theorems, combining ingredients from the previous sections.

Theorem 5.8.4. *Suppose that $w \in I$ and that $f \in H^0(X_1^{G_1,I},\omega^2(-D))$ satisfies $U_{\mathrm{Kli}(w),1}f = (\alpha_w + \beta_w)f$, $U_{\mathrm{Iw}(w),1}f = \alpha_w f$, and $U_{w,2}f = \alpha_w\beta_w f$, where $\alpha_w, \beta_w \neq 0$. Then $f = 0$.*

Proof. First suppose that the restriction of f to $X_1^{G_1,I,=_w 1}$ is zero. Then we may write $f = \mathrm{Ha}_w g$ for some $g \in H^0(X_1^{G_1,I},\omega^2 \otimes \omega_w^{1-p}(-D))$. Moreover because we have $Z_w f = \beta_w f$ by hypothesis, we would then have $Z_w g = \beta_w g$ by Proposition 5.4.2. But then by Corollary 5.4.5, $g = 0$ and hence $f = 0$.

Now we may suppose that the restriction of f to $X_1^{G_1,I,=_w 1}$ is nonzero. Combining Proposition 5.6.7 with Proposition 5.7.9 there is a commutative diagram

$$\begin{tikzcd}
H^0(Y_1^{G_1,I},\omega^2) \arrow[r] \arrow[d, "\mathrm{Ha}_w U_{\mathrm{Iw}(w),1}"] \arrow[dd, bend right=60, "U_{w,2}"'] & H^0(Y_1^{G_1,I,=_w 1},\omega^2|_{Y_1^{G_1,I,=_w 1}}) \arrow[d, "\gamma_1"] \\
H^0(Y_1^{G_1,I},\omega^2 \otimes \omega_w^{p-1}) \arrow[r] & H^0(Y_1^{G_1,I,=_w 1},\omega^2 \otimes \omega_w^{p-1}|_{Y_1^{G_1,I,=_w 1}}) \arrow[d, "\gamma_2"] \\
H^0(Y_1^{G_1,I},\omega^2) \arrow[r] & H^0(Y_1^{G_1,I,=_w 1},\omega^2|_{Y_1^{G_1,I,=_w 1}})
\end{tikzcd}$$

where the horizontal maps are the natural restriction maps, and the vertical maps on the right column are as in diagrams (5.6.6) and (5.7.7).

If we start with f in the top left of the diagram, we obtain something nonzero on the bottom right because $U_{w,2}f = \alpha_w\beta_w f$ and the restriction of f to $Y_1^{G_1,I,=_w 1}$ is nonzero. The commutativity of the diagram implies that $\mathrm{Ha}_w U_{\mathrm{Iw}(w),1}f$ has nonzero restriction to $Y_1^{G_1,I,=_w 1}$. On the other hand, because $U_{\mathrm{Iw}(w),1}f = \alpha_w f$, $\mathrm{Ha}_w U_{\mathrm{Iw}(w),1}f = \alpha_w \mathrm{Ha}_w f$ which vanishes along $Y_1^{G_1,I,=_w 1}$. This is a contradiction. □

Remark 5.8.5. In fact, something stronger than Theorem 5.8.4 is true: if $w \in I$, and $f \in H^0(X_1^{G_1,I},\omega^2(-D))$ satisfies $U_{w,2}f \neq 0$, then $U_{\mathrm{Iw}(w),1}f \notin H^0(X_1^{G_1,I},\omega^2(-D))$. This can be proved in exactly the same way as Theorem 5.8.4, given the following strengthening of Corollary 5.4.5: if $w \in I$, then

$$H^0(X_{K_p(I)K^p,1}^{G_1,I},\omega^2 \otimes \omega_w^{1-p}(-D)) = 0.$$

In the case $p > 3$, we will sketch a proof of this result in §5.9 using Fourier–Jacobi expansions, but since a complete argument in the case $p = 3$ would involve developing considerably more of the details of toroidal compactifications than we need in the rest of the paper, we have decided not to give the details.

When $p > 3$, this vanishing result holds even for non cusp forms, so the same is true of Theorem 5.8.4.

We can now prove the injectivity of the doubling map.

Theorem 5.8.6 (Doubling)**.** *The doubling map*

$$\begin{aligned} e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D)) \oplus e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D)) \\ \to e(U_{\mathrm{Iw}(w),1}U_{w,2})H^0(X_1^{G_1,I,=_w 2},\omega^2(-D)) \end{aligned}$$

is injective.

Proof. Assume the map is not injective. By Lemma 5.8.2, the kernel is an inductive limit of finite dimensional vector spaces stable under the commuting operators $U_{w,2}$ and $U_{\mathrm{Iw}(w),1}$. We may therefore take a nonzero simultaneous eigenvector (f, g) for $U_{w,2}$ and $U_{\mathrm{Iw}(w),1}$ in this kernel, with respective eigenvalues $\alpha_w\beta_w$ and β_w for some α_w, $\beta_w \neq 0$ (the eigenvalues are nonzero because we are by assumption in the ordinary space for both $U_{\mathrm{Iw}(w),1}$ and $U_{w,2}$). It follows from the definition of the action of $U_{\mathrm{Iw}(w),1}$ that $f = -\beta_w g$ and $U_{\mathrm{Kli}(w),1}f = (\alpha_w + \beta_w)f$. Since we are also assuming that $f + Z_w g = 0$, we see that the image of f in $H^0(X_1^{G_1,I,=_w 2},\omega^2(-D))$ satisfies $U_{\mathrm{Iw}(w),1}f = \alpha_w f$. The result follows from Theorem 5.8.4 (note that the eigenvalues for $U_{\mathrm{Iw}(w),1}$ and $U_{w,2}$ are nonzero because we are in the ordinary space for these operators by hypothesis). □

Remark 5.8.7. We now put the Theorem 5.8.6 in a form that is used in §7.9. Assume that $M \subset e(U_{w,2})H^0(X_1^{G_1,I},\omega^2(-D)) \otimes \bar{\mathbf{F}}_p$ is a finite dimensional vector space, stable under $U_{\mathrm{Kli}(w),1}$ and $U_{w,2}$. Assume that there are distinct elements $\alpha_w, \beta_w \in \bar{\mathbf{F}}_p^\times$ such that $U_{\mathrm{Kli}(w),1} - (\alpha_w + \beta_w)$ and $U_{w,2} - \alpha_w\beta_w$ are nilpotent on M. The sub-algebra $\mathcal{E}$ of $\mathrm{End}(M)$ generated by $U_{\mathrm{Kli}(w),1}$ and $U_{w,2}$ is therefore a local Artinian algebra and there are elements $\tilde{\alpha}_w, \tilde{\beta}_w \in \mathcal{E}$ satisfying $\tilde{\alpha}_w = \alpha_w \mod \mathfrak{m}_{\mathcal{E}}$ and $\tilde{\beta}_w = \beta_w \mod \mathfrak{m}_{\mathcal{E}}$ and such that on $M \oplus M$ we have $(U_{\mathrm{Iw}(w),1} - \tilde{\alpha}_w)(U_{\mathrm{Iw}(w),1} - \tilde{\beta}_w) = 0$.

We can define maps $\iota_{\xi_w} : M \to M \oplus M$ by $f \mapsto (f, -\tilde{\xi}_w^{-1} f)$ for $\xi_w \in \{\alpha_w, \beta_w\}$. Then one checks easily that the map $\iota_{\alpha_w} \oplus \iota_{\beta_w} : M \oplus M \to M \oplus M$ is an isomorphism and that the composite with the doubling map takes the form $(f_1, f_2) \mapsto ((1 - \tilde{\beta}_w U_{\mathrm{Iw}(w),1}^{-1})f_1 + (1 - \tilde{\alpha}_w U_{\mathrm{Iw}(w),1}^{-1})f_2)$. The first and second components of this map therefore define injective maps

$$M \hookrightarrow e(U_{\mathrm{Iw}(w),1}U_{w,2})H^0(X_1^{G_1,I,=_w 2},\omega^2(-D))_{U_{\mathrm{Iw},1}-\xi_w}.$$

for ξ_w respectively equal to α_w and β_w.

5.9. **Vanishing in partial negative weight: Fourier–Jacobi expansions.** We end this section by giving a proof of the following vanishing result in "partial negative weight", which partially strengthens Corollary 5.4.5 but is not needed in this paper (see also Remark 5.8.5).

Proposition 5.9.1. *Assume $w \in I$. If $p > 3$ and $[F : \mathbf{Q}] > 1$, then*

$$H^0(X^I_{K_p(I)K^p,1}, \omega^2 \otimes \omega_w^{1-p}) = 0.$$

Remark 5.9.2. We have a sketch of an argument to show that if $p = 3$, then $H^0(X^I_{K_p(I)K^p,1}, \omega^2 \otimes \omega_w^{1-p}(-D)) = 0$. We also have a sketch of an argument for $F = \mathbf{Q}$. But to give complete proofs would require us to justify certain properties of Fourier–Jacobi expansion for which we could not find references (for example we would need to have good geometric theory of cuspidal Fourier–Jacobi forms).

We will prove Proposition 5.9.1 by restriction to a boundary stratum, and ultimately reducing to the vanishing of spaces of Hilbert modular forms of partial negative weight.

We let $K_p = \prod_{v|p} \mathrm{GSp}_4(\mathcal{O}_{F_v})$, and by possibly shrinking K^p we may assume that it is a principal level structure in the sense of [Lan13, §1.3.6]. We let $c \in \mathbf{Z}^{\times,+}_{(p)} \backslash (\mathbf{A}^{\infty,p} \otimes F)^\times / \nu(K^p)$, and we may work with the connected component $X_{K,1,c}$ of $X_{K,1}$. We now choose a boundary stratum $Z \hookrightarrow X_{K,1,c}$ corresponding to a one dimensional totally isotropic factor $W \in \mathfrak{C}$ (see §3.5). It means that the restriction of the semi-abelian scheme along Z is an extension of an abelian scheme A of dimension $[F : \mathbf{Q}]$ with $\mathcal{O}_F$-action by a torus T of dimension $[F : \mathbf{Q}]$ with $\mathcal{O}_F$-action.

Let $H := \ker(\mathrm{Res}_{F/\mathbf{Q}} \mathrm{GL}_2 \to (\mathrm{Res}_{F/\mathbf{Q}} \mathbf{G}_m)/\mathbf{G}_m)$. The abelian scheme of dimension $[F : \mathbf{Q}]$ is parametrized by a (connected) Shimura variety for the group H (this is a Hilbert–Blumenthal modular variety) that we denote by $Y_{H,1}$ and is a moduli space of isomorphism classes of triples $(A, \iota, \lambda, \eta)$:

(1) $A \to \operatorname{Spec} R$ is an abelian scheme,
(2) $\iota : \mathcal{O}_F \to \mathrm{End}(G) \otimes \mathbf{Z}_{(p)}$ is an action,
(3) $\mathrm{Lie}(A)$ is a locally free $\mathcal{O}_F \otimes_{\mathbf{Z}} R$-module of rank 1,
(4) $\lambda : A \to A^t$ is a prime to p, $\mathcal{O}_F$-linear quasi-polarization such that for all $v|p$, $\mathrm{Ker}(\lambda : A[v^\infty] \to A^t[v^\infty])$ is trivial,
(5) η is a prime to p level structure.

We denote by $X_{H,1}$ a toroidal compactification of $Y_{H,1}$. We have partial Hasse invariants A_v for all $v|p$. Let $Y^I_{H,1} \subset X^I_{H,1}$ be the Zariski opens where A_v is invertible for all $v \in I^c$. We have a map $Z \to Y_{H,1}$ and we let $Z^I = Z \times_{Y_{H,1}} Y^I_{H,1}$.

The étale map $X_{K_p(I)K^p,1} \to X_{K,1}$ has a section along $Z^I \hookrightarrow X_{K,1}$ which is provided by the rank one multiplicative groups $T[v]$ for all $v \in I$. Therefore the map $X_{K_p(I)K^p,1} \to X_{K,1}$ has a section restricted to the completion of $X_{K,1}$ along Z^I.

Proposition 5.9.3 (Fourier–Jacobi expansion principle)**.** *There is a natural injective Fourier–Jacobi expansion map*

$$H^0(X^I_{K_p(I)K^p,1,c}, \omega^2 \otimes \omega_v^{1-p}) \to \prod_{\xi \in \mathfrak{a}^+} H^0(A^I_1, \omega^2 \otimes \omega_v^{1-p} \otimes \mathcal{L}_\xi)$$

where $\mathfrak{a}$ is a fractional ideal of $\mathcal{O}_F$ and $\mathfrak{a}^+$ are the positive elements, $A^I_1 \to Y^I_{H,1}$ is an abelian scheme isogenous to the universal abelian scheme A and $\mathcal{L}_\xi$ is an invertible sheaf over A^I_1, rigidified along the identity section.

Proof. The existence of such a map follows from the description of the toroidal boundary charts, as in [FC90, §V] or [Lan13, §6.2.3,§7.1]. It is obtained by restricting sections to the completion along Z^I. The sheaves $\mathcal{L}_\xi$ are obtained by pullback from a Poincaré bundle which is rigidified along the identity section.

The injectivity result is clear as long as we can show that $X^I_{K_p(I)K^p,1,c}$ is connected. This follows directly from the connectedness of $X_{K,1,c}$ and the irreducibility of the Igusa tower, for which see [Hid04, Cor. 8.17] or [Hid09, Thm. 0.1]. □

We will now prove the vanishing of the groups $H^0(A^I_1, \omega^2 \otimes \omega_w^{1-p} \otimes \mathcal{L}_\xi)$. We first need the following preliminary lemma.

Lemma 5.9.4. *Let S be a scheme and let $A \to S$ be an abelian scheme. Let $\mathcal{L}$ be an invertible sheaf on A, rigidified along the unit section. Then for all $n \in \mathbf{Z}_{\geq 1}$, $\mathcal{L}^{n^3}|_{A[n]}$ is trivial.*

Proof. It is well-known ([Mum08, Chap. II, §6 and §8]) that $n^*\mathcal{L} \simeq \mathcal{L}^{n^2} \otimes \mathcal{L}_0$ where $\mathcal{L}_0$ is a sheaf algebraically equivalent to zero. Moreover $n^*\mathcal{L}_0 \simeq \mathcal{L}_0^n$. Therefore $\mathcal{L}^{n^3} \simeq (n^*\mathcal{L})^n \otimes n^*\mathcal{L}_0^{-1}$ is trivial on $A[n]$. □

Now we may prove the following sequence of vanishing results for negative weight forms. (Note that the first part is a very special case of the main theorem of [DK17], although the argument there is different.)

Proposition 5.9.5. *Assume that $[F : \mathbf{Q}] > 1$. Let $\kappa = (k_v)_{v \in S_p}$ be a weight for H and suppose that there is a $w \in I$ such that $k_w < 0$. Then:*

(1) $H^0(Y_{H,1}, \omega^\kappa) = 0$.
(2) $H^0(Y^I_{H,1}, \omega^\kappa) = 0$.
(3) For any $\xi \in \mathfrak{a}^+$, $H^0(A^I_1, \omega^\kappa \otimes \mathcal{L}_\xi) = 0$.

Proof. We derive each claim in turn from the previous one:

(1) Let $C_w \subset Y_{H,1}$ be the simultaneous vanishing locus of the Hasse invariants A_v for $v \neq w$; it is a (union of) smooth curves (since p is split completely, this is an easy local calculation). Furthermore, because $[F : \mathbf{Q}] > 1$, it is also proper (note that if $[F : \mathbf{Q}] = 1$, then $C_w = Y_{H,1}$ is not proper).

By the existence of the secondary Hasse invariants, $\omega_v|_{C_w}$ is a torsion line bundle for $v \neq w$, while $\omega_w|_{C_w}$ has positive degree on each component. Let $\mathcal{I}$ be the ideal sheaf of C_w in $Y_{H,1}$. It follows from the Kodaira–Spencer isomorphism that we have an isomorphism

$$\mathcal{I}/\mathcal{I}^2 = \bigoplus_{v \neq w} \omega_v^2.$$

Thus for all $m \geq 0$, $\mathcal{I}^m/\mathcal{I}^{m+1} = \mathrm{Sym}^m \mathcal{I}/\mathcal{I}^2$ is a direct sum of torsion line bundles. Because $k_w < 0$, it follows that $\mathcal{I}^m/\mathcal{I}^{m+1} \otimes \omega^\kappa$ is a direct sum of line bundles of negative degree, and hence has no sections. The result follows from this and the fact that every irreducible component of $X_{H,1}$ contains a component of C_w (by considering the formal expansion of any form along C_w).

(2) If $f \in H^0(Y^I_{H,1}, \omega^\kappa)$, then, for $c_v \gg 0$ for all $v \in I^c$,

$$f \prod_{v \in I^c} A_v^{c_v} \in H^0(Y_{H,1}, \omega^\kappa \otimes \bigotimes_{v \in I^c} \omega_v^{c_v(p-1)})$$

and hence vanishes by part (1). Thus the same conclusion holds for f.

(3) For all $n \in \mathbf{Z}_{\geq 1}$ with $(n, p) = 1$ we will show that any section of $f \in H^0(A^I_1, \omega^\kappa \otimes \mathcal{L}_\xi)$ vanishes on the n-torsion subgroup $A^I_1[n]$, and hence vanishes identically. After replacing f by f^{n^3}, we can assume that $\mathcal{L}_\xi$ is the

trivial sheaf (see Lemma 5.9.4). We then consider the norm of f for each irreducible component of the finite étale map $A_1^I[n] \to Y_{H,1}^I$ to reduce to part (2). □

Proof of Prop. 5.9.1. This is an immediate consequence of Proposition 5.9.3 and Proposition 5.9.5 (3), because all the terms in the Fourier–Jacobi expansion will be zero. □

6. Higher Coleman theory

In this section, we construct (higher) Coleman theories for $\mathrm{GSp}_4(\mathbf{A}_F)$. As in §4, we assume that p splits completely in F and we construct all possible Coleman theories, allowing the weight space at each place above p to be either one or two-dimensional. In the case that $F = \mathbf{Q}$ this was carried out in [AIP15] and [Pil20]. Many of our arguments are simply the "product over the places $v|p$" of the arguments of [AIP15] and [Pil20]. To keep this paper at a reasonable length, we will often refer to these papers for the details of arguments which go over directly to our case.

The main results of this section are Theorem 6.5.8 (a classicality result for overconvergent cohomology classes of small slope), and Theorem 6.6.4 (which shows that in the case that I has size at most one, the cohomology of the Hida complex M_I constructed in §4 is overconvergent, once p is inverted). These results together improve (at the expense of inverting p) on the classicality results of §4, in that they do not require the weight to be sufficiently large; this is crucial for our applications to abelian surfaces, which correspond to modular forms of parallel weight 2.

We begin in §6.1 with the construction at the level of formal schemes of a version of the analytic sheaves of overconvergent forms that we will use later in this section. The purpose of these sheaves is to allow us in §6.2 to show that the cohomology of our analytic complexes is concentrated in degrees $[0, \#I]$; as usual, this involves a comparison of the toroidal and minimal compactifications, and we do not know how to carry out this argument purely in the analytic setting. In §6.3 we construct the corresponding structures in the analytic world, and we show that an appropriate Hecke operator (a product of "U_p" operators at the places dividing p) acts compactly.

We then recall in §6.4 the analytic BGG resolution comparing the cohomology with locally analytic coefficients to that with algebraic coefficients, which is one of the ingredients in our small slope classicality theorem, which is proved in §6.5, the other ingredient being a version of the analytic continuation argument of [Kas06]. Finally, in §6.6 we apply our results to the complexes constructed in §4.6. We are only able to show that the ordinary cohomology is overconvergent if $\#I \leq 1$; fortunately, this suffices for the arguments that we make in §7.

6.1. Sheaves of overconvergent and locally analytic modular forms: the formal construction.

In this section the base is $\mathbf{C}_p$, the p-adic completion of an algebraic closure of $\mathbf{Q}_p$. We will construct overconvergent versions of our interpolation sheaves Ω^{κ_I} and develop a finite slope theory. It is necessary to connect the ordinary theory and the slope 0 overconvergent theory, because we are only able to prove a strong classicality theorem in the overconvergent setting. In the first part of this section, we begin by working at a formal level. The reason is that we need

to prove a vanishing theorem (Theorem 6.2.6) for the overconvergent cohomology and we don't know how to prove it without using formal models.

6.1.1. *Slope decompositions.* We very briefly recall the basics of the theory of slope decompositions for compact operators, which was introduced in [AS08] and further developed in [Urb11]. Given a vector space M over $\mathbf{C}_p$ with a linear endomorphism U, and a rational number h, an h-slope decomposition of M with respect to U is a decomposition $M = M^{\leq h} \oplus M^{>h}$ into U-stable subspaces, where

- $M^{\leq h}$ is finite-dimensional,
- all of the eigenvalues a of U on $M^{\leq h}$ have $v(a) \leq h$, and
- if Q is a monic polynomial whose roots all have valuation less than h, then $Q(U)$ acts invertibly on $M^{>h}$.

If slope decompositions exist, they are unique. If they exist for all h, then we say that the finite slope part is the union of the $M^{\leq h}$ for all $h \in \mathbf{Q}$.

The notion of a slope decomposition can be generalized to the case of modules over a $\mathbf{C}_p$-Banach algebra A. In particular, it is known that compact operators on projective A-Banach modules admit slope decompositions locally on $\operatorname{Max} A$. It is explained in [Urb11, §2] and [Pil20, §13] how to generalize this notion to perfect complexes of modules over Banach algebras. In brief, an endomorphism U of a perfect complex is said to be compact if it admits a representative $\widetilde{U}$ as an endomorphism of a bounded complex $M^\bullet$ of projective Banach modules, which is compact in each degree. Then one may consider the product of characteristic power series of $\widetilde{U}$ on the individual M^i, and the corresponding spectral variety for $\widetilde{U}$ as in [Col97]. The complex $M^\bullet$ determines a complex of coherent sheaves $\mathcal{M}^\bullet$ over this spectral variety, and one defines the spectral variety of U to be the support of the cohomology sheaves $H^\bullet(\mathcal{M}^\bullet)$. One checks that this is independent of the choice of $M^\bullet$ and $\widetilde{U}$. The sheaves $H^\bullet(\mathcal{M}^\bullet)$ over the spectral variety for U admit slope decompositions.

6.1.2. *Recollections about formal Banach sheaves.* An *admissible* $\mathcal{O}_{\mathbf{C}_p}$-algebra is a flat $\mathcal{O}_{\mathbf{C}_p}$-algebra which is a quotient of a converging power series ring $\mathcal{O}_{\mathbf{C}_p}\langle X_1, \cdots, X_n\rangle$ by a finitely generated ideal. In this section we work with quasi-compact and separated p-adic formal schemes over Spf $\mathcal{O}_{\mathbf{C}_p}$ which admit an open covering by formal spectra of admissible algebras. We call these formal schemes admissible. (In some parts of the literature, an admissible affine formal scheme Spf A is one for which A is admissible, in the sense that it is a complete and separated topological ring, which is linearly topologized and has an ideal of definition, i.e. an open ideal I such that every neighbourhood of 0 contains some power of I. Our admissible algebras are a special case of this definition, and we hope that our terminology will not cause any confusion.)

We recall some definitions taken from [AIP15, Defn. A.1.1.1]. We let $\mathfrak{S}$ be an admissible formal scheme. A formal Banach sheaf over $\mathfrak{S}$ is a family $(\mathfrak{F}_n)_{n\geq 0}$ of quasi-coherent sheaves such that:

(1) $\mathfrak{F}_n$ is a sheaf of $\mathcal{O}_{\mathfrak{S}}/p^n$-modules,
(2) $\mathfrak{F}_n$ is flat over $\mathcal{O}_{\mathbf{C}_p}/p^n$,
(3) For all $0 \leq m \leq n$, we have isomorphisms $\mathfrak{F}_n \otimes_{\mathcal{O}_{\mathbf{C}_p}} \mathcal{O}_{\mathbf{C}_p}/p^m \simeq \mathfrak{F}_m$.

We can associate to $(\mathfrak{F}_n)_n$ a sheaf $\mathfrak{F}$ over $\mathfrak{S}$ equal to the inverse limit $\varprojlim_n \mathfrak{F}_n$ (the maps in the inverse limit are those provided by (3) above). Since $\mathfrak{F}_n = \mathfrak{F} \otimes_{\mathcal{O}_{\mathbf{C}_p}}$

$\mathcal{O}_{\mathbf{C}_p}/p^n$, the sheaf $\mathfrak{F}$ clearly determines the $(\mathfrak{F}_n)$ and we identify $\mathfrak{F}$ and the family $(\mathfrak{F}_n)$ in the sequel. We say that a Banach sheaf $\mathfrak{F}$ is flat if $\mathfrak{F}_n$ is a flat $\mathcal{O}_{\mathfrak{S}}/p^n$-module for all n.

We say that a Banach sheaf $\mathfrak{F}$ is small if there exists a coherent $\mathcal{O}_{\mathfrak{S}}/p$-module $\mathcal{F}$ such that $\mathfrak{F}_1$ is an inductive limit of coherent sheaves $\varinjlim_{j\in\mathbf{N}} \mathfrak{F}_{1,j}$ and the quotients $\mathfrak{F}_{1,j+1}/\mathfrak{F}_{1,j}$ are direct summands of $\mathcal{F}$.

We now recall a vanishing result from [AIP15].

Theorem 6.1.3. *Let $\mathfrak{S}$ be an admissible formal scheme. Assume that $\mathfrak{S}$ admits a projective map $\mathfrak{S} \to \mathfrak{S}'$ to an affine admissible formal scheme which induces an isomorphism of the associated analytic adic spaces over* $\mathrm{Spa}(\mathbf{C}_p, \mathcal{O}_{\mathbf{C}_p})$. *Let $\mathfrak{F}$ be a small Banach sheaf over $\mathfrak{S}$. Let $\mathfrak{U}$ be an affine cover of $\mathfrak{S}$. Then the Čech complex*

$$Cech(\mathfrak{U}, \mathfrak{F}) \otimes_{\mathcal{O}_{\mathbf{C}_p}} \mathbf{C}_p$$

is acyclic in positive degree.

Proof. This is a special case of [AIP15, Thm. A.1.2.2]. Indeed, the proof of [AIP15, Thm. A.1.2.2] is by reducing to this case, which is case (1) of that proof. □

6.1.4. *Recollections about the Hodge–Tate period map.* If $H \to \operatorname{Spec} S$ is a finite flat group scheme, we denote by H^D its Cartier dual and by ω_{H^D} the conormal sheaf of H^D along its unit section. This is a coherent $\mathcal{O}_S$-module. We can view ω_{H^D} as an *fppf*-sheaf of abelian groups. If $q : T \to S$ is an S-scheme, we let $\omega_{H^D}(T) = H^0(T, q^*\omega_{H^D})$. There is a well-known Hodge–Tate map $\mathrm{HT}_H : H \to \omega_{H^D}$ of *fppf*-sheaves of abelian groups which associates to any S-scheme T and point $x \in H(T)$ the differential $x^*\frac{\mathrm{d}t}{t}$, where we are (thanks to Cartier duality) viewing x as a morphism $x : H^D_T \to \mathbf{G}_m|_T$ of T-group schemes.

Let $K = K_pK^p$ be a neat compact open subgroup with $K_p = \prod_{v|p} \mathrm{GSp}_4(\mathcal{O}_{F_v})$. Consider the non-compactified Shimura variety $Y_K \to \operatorname{Spec} \mathcal{O}_{\mathbf{C}_p}$. We denote by $\mathfrak{Y}_K \to \mathrm{Spf}\ \mathcal{O}_{\mathbf{C}_p}$ the associated p-adic formal scheme. We fix a toroidal compactification $Y_K \hookrightarrow X_K$ and denote by $\mathfrak{X}_K$ the p-adic formal scheme associated to X_K. Let $\mathcal{Y}_K \hookrightarrow \mathcal{X}_K$ be the associated analytic adic spaces over $\mathrm{Spa}\ (\mathbf{C}_p, \mathcal{O}_{\mathbf{C}_p})$.

Let $n = (n_v)_{v\in S_p} \in \mathbf{Z}^{S_p}_{\geq 0}$. We let $K(p^n)$ be the compact open subgroup defined by $K(p^n) = K_p(p^n)K^p$ where $K_p(p^n) = \prod_v \mathrm{Ker}\big(\mathrm{GSp}_4(\mathcal{O}_{F_v}) \to \mathrm{GSp}_4(\mathcal{O}_{F_v}/p^{n_v})\big)$ is the principal congruence subgroup of level n.

We let $Y_{K(p^n),\mathbf{C}_p} \to Y_K \times_{\operatorname{Spec}\mathcal{O}_{\mathbf{C}_p}} \operatorname{Spec} \mathbf{C}_p$ be the Shimura variety with level $K(p^n)$ structure over $\operatorname{Spec} \mathbf{C}_p$. This map is finite étale with Galois group equal to $\prod_{v|p} \mathrm{GSp}_4(\mathcal{O}_{F_v}/p^{n_v})$. Associated to our choice of polyhedral cone decomposition we have a toroidal compactification $Y_{K(p^n),\mathbf{C}_p} \to X_{K(p^n),\mathbf{C}_p}$. We denote by $\mathcal{Y}_{K(p^n)} \hookrightarrow \mathcal{X}_{K(p^n)}$ the associated analytic spaces over $\mathrm{Spa}(\mathbf{C}_p, \mathcal{O}_{\mathbf{C}_p})$. The map $\mathcal{X}_{K(p^n)} \to \mathcal{X}_K$ is finite flat. We denote by $\mathfrak{X}_{K(p^n)} \to \mathfrak{X}_K$ the normalization of $\mathfrak{X}_K$ in $\mathcal{X}_{K(p^n)}$ and by $\mathfrak{Y}_{K(p^n)}$ the normalization of $\mathfrak{Y}_K$ in $\mathcal{Y}_{K(p^n)}$. These are admissible formal schemes (see [PS16a, §1.1]). There is a universal, $\mathcal{O}_F$-linear map $\prod_{v|p}(\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_F)^4 \to \prod_{v|p} \mathcal{G}_v[p^{n_v}]$ over $\mathfrak{Y}_{K(p^n)}$, which is symplectic up to a similitude factor and is an isomorphism on the associated analytic adic spaces.

There is a Hodge–Tate period map $\mathrm{HT} : \prod_{v|p} \mathcal{G}_v[p^{n_v}] \to \prod_{v|p} \omega_{\mathcal{G}_v}/p^{n_v}\omega_{\mathcal{G}_v}$ (we are using the quasi-polarization of $\mathcal{G}_v$ to identify $\mathcal{G}_v$ and $\mathcal{G}_v^D$) which we can compose with $\prod_{v|p}(\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4 \to \prod_{v|p} \mathcal{G}_v[p^{n_v}]$ to obtain an $\mathcal{O}_F$-linear map of sheaves

over $\mathfrak{Y}_{K(p^n)}$

$$\mathrm{HT} : \prod_{v|p} (\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4 \to \prod_{v|p} \omega_{\mathcal{G}_v}/p^{n_v}\omega_{\mathcal{G}_v}.$$

We claim that this map admits an extension

$$\mathrm{HT} : \prod_{v|p} (\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4 \to \prod_{v|p} \omega_{\mathcal{G}_v}/p^{n_v}\omega_{\mathcal{G}_v}$$

over $\mathfrak{X}_{K(p^n)}$. When $F = \mathbf{Q}$, this is the content of [PS16a, Prop. 1.2]. For a general F we can use the Koecher principle of [Lan17, Thm. 8.7].

According to a result of Fargues ([Far10, Thm. 7], see also [PS16a, Thm. 1.5]), the cokernel of the linearization of HT is annihilated by $p^{\frac{1}{p-1}}$. By [PS16a, §1.4] when $F = \mathbf{Q}$ (and an immediate generalization for general F), there exists an admissible formal scheme $\mathfrak{X}^{mod}_{K(p^n)} \to \mathfrak{X}_{K(p^n)}$, which is the normalization of a blow-up (the ideal of the blow-up is finitely generated and contains a power of p), and a modification $\omega^{mod}_{\mathcal{G}} \subset \omega_{\mathcal{G}}$ such that:

(1) $\omega^{mod}_{\mathcal{G}}$ is a locally free $\mathcal{O}_F \otimes \mathcal{O}_{\mathfrak{X}^{mod}_{K(p^n)}}$-module of rank 2,
(2) $p^{\frac{1}{p-1}}\omega_{\mathcal{G}} \subset \omega^{mod}_{\mathcal{G}} \subset \omega_{\mathcal{G}}$,
(3) The Hodge–Tate map HT factorizes into a map

$$\mathrm{HT} : \prod_{v|p} (\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4 \to \prod_{v|p} \omega^{mod}_{\mathcal{G}_v}/p^{n_v - \frac{1}{p-1}}\omega^{mod}_{\mathcal{G}_v}$$

and the linearized map

$$\mathrm{HT} \otimes 1 : \prod_{v|p} (\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4 \otimes \mathcal{O}_{\mathfrak{X}^{mod}_{K(p^n)}} \to \prod_{v|p} \omega^{mod}_{\mathcal{G}_v}/p^{n_v - \frac{1}{p-1}}\omega^{mod}_{\mathcal{G}_v}$$

is surjective.

We say a few words about the construction of this formal model. We first introduce the subsheaf $\omega^{mod}_{\mathcal{G}}$ of $\omega_{\mathcal{G}}$ generated over $\mathfrak{X}_{K(p^n)}$ by $p^{\frac{1}{p-1}}\omega_{\mathcal{G}}$ and local lifts of $\mathrm{HT}(\prod_{v|p}(\mathcal{O}_{F_v}/p^{n_v}\mathcal{O}_{F_v})^4)$ in $\omega_{\mathcal{G}}$. The sheaf $\omega^{mod}_{\mathcal{G}}$ constructed in this way is not locally free, but becomes locally free after pulling back to $\mathfrak{X}^{mod}_{K(p^n)}$ (and we continue to denote this pulled back sheaf by $\omega^{mod}_{\mathcal{G}}$). We now describe the procedure used to construct $\mathfrak{X}^{mod}_{K(p^n)}$. Zariski locally over $\mathfrak{X}_{K(p^n)}$ we can find a map

$$\prod_{v|p} (\mathcal{O}_{F_v} \otimes_{\mathbf{Z}_p} \mathcal{O}_{\mathfrak{X}_{K(p^n)}})^4 \to \omega_{\mathcal{G}}$$

by considering local lifts of the Hodge-Tate classes in $\omega_{\mathcal{G}}$, and the image of this map is $\omega^{mod}_{\mathcal{G}}$. Zariski locally, we can trivialize $\omega_{\mathcal{G}}$ and we can represent the above map by a 2×4 matrix at each place $v|p$. The formal scheme is obtained by taking the normalization of the blow-up of the ideal which is the product at all places v dividing p of the ideal locally generated by the 2×2-minors of the matrix at v.

We denote by $e_{v,1}, \ldots, e_{v,4}$ the canonical basis of $\mathcal{O}^4_{F_v}$. We let $\epsilon = (\epsilon_v)_{v|p} \in \prod_{v|p}([0, n_v - \frac{1}{p-1}] \cap \mathbf{Q})$. We define an admissible formal scheme $\mathfrak{X}_{K(p^n)}(\epsilon) \to \mathfrak{X}^{mod}_{K(p^n)}$ (an open subscheme of an admissible blowup of $\mathfrak{X}^{mod}_{K(p^n)}$) by the conditions that:

- $\mathrm{HT}(e_{v,1}) \in p^{\epsilon_v}\omega^{mod}_{\mathcal{G}_v}/p^{n_v - \frac{1}{p-1}}\omega^{mod}_{\mathcal{G}_v}$ for all $v|p$,
- $\mathrm{HT}(e_{v,2}) \in p^{\epsilon_v}\omega^{mod}_{\mathcal{G}_v}/p^{n_v - \frac{1}{p-1}}\omega^{mod}_{\mathcal{G}_v}$ for all $v \in I^c$.

For all $v \in I^c$, the Hodge–Tate map factorizes into an isomorphism $\mathrm{HT} \otimes 1 : \mathcal{O}_{\mathfrak{X}_{K(p^n)}}(\epsilon)/p^{\epsilon_v} e_{v,3} \oplus \mathcal{O}_{\mathfrak{X}_{K(p^n)}}(\epsilon)/p^{\epsilon_v} e_{v,4} \to \omega_{\mathcal{G}_v}^{mod}/p^{\epsilon_v}\omega_{\mathcal{G}_v}^{mod}$.

For all $v|p$, we let $\mathrm{Fil}_v^{\mathrm{can}} \subset \omega_{\mathcal{G}_v}^{mod}/p^{\epsilon_v}$ be the sub-module generated by $\mathrm{HT}(e_{v,2})$ and $\mathrm{HT}(e_{v,3})$.

Lemma 6.1.5. $\mathrm{Fil}_v^{\mathrm{can}}$ *is a locally free* $\mathcal{O}_{\mathfrak{X}_{K(p^n)}}(\epsilon)/p^{\epsilon_v}$*- module of rank one, and is locally a direct factor in* $\omega_{\mathcal{G}_v}^{mod}/p^{\epsilon_v}\omega_{\mathcal{G}_v}^{mod}$.

Proof. See [Pil20, Lem. 12.2.2.1]. □

We let $\mathrm{Gr}_v^{\mathrm{can}} = \omega_{\mathcal{G}_v}^{mod}/(p^{\epsilon_v}\omega_{\mathcal{G}_v}^{mod} + \mathrm{Fil}_v^{\mathrm{can}})$. Then for all $v|p$, the Hodge–Tate map induces an isomorphism:

$$\mathrm{HT} \otimes 1 : (\mathcal{O}_{\mathfrak{X}_{K(p^n)}}(\epsilon)/p^{\epsilon_v})e_{v,4} \to \mathrm{Gr}_v^{\mathrm{can}}.$$

If $v \in I^c$, the Hodge–Tate map also induces an isomorphism

$$\mathrm{HT} \otimes 1 : (\mathcal{O}_{\mathfrak{X}_{K(p^n)}}(\epsilon)/p^{\epsilon_v})e_{v,3} \to \mathrm{Fil}_v^{\mathrm{can}}.$$

6.1.6. *Flag varieties.* We denote by $\mathfrak{F}\mathfrak{L}_n \to \mathfrak{X}_{K(p^n)}^{mod}$ the flag formal scheme which parametrizes locally free direct summands of rank one (as $\mathcal{O}_F \otimes \mathcal{O}_{\mathfrak{X}_{K(p^n)}^{mod}}$-modules) $\mathrm{Fil}\omega_{\mathcal{G}}^{mod}$ in $\omega_{\mathcal{G}}^{mod}$. This space decomposes into a product $\mathfrak{F}\mathfrak{L}_n = \prod_{v|p} \mathfrak{F}\mathfrak{L}_{v,n}$ over all places v above p.

Let $w = (w_v) \in \prod_{v|p}[0, \epsilon_v] \cap \mathbf{Q}$. We let $\mathfrak{F}\mathfrak{L}_{n,\epsilon,w} \to \mathfrak{X}_{K(p^n)}(\epsilon)$ be the moduli space of locally free direct summands of rank one $\mathrm{Fil}\omega_{\mathcal{G}_v}^{mod} \subset \omega_{\mathcal{G}_v}^{mod}$ such that $\mathrm{Fil}\omega_{\mathcal{G}_v}^{mod} = \mathrm{Fil}_v^{\mathrm{can}} \mod p^{w_v}$.

We let $w' = (w'_v) \in \prod_{v|p}[0, w_v] \cap \mathbf{Q}$. We let $\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'} \to \mathfrak{F}\mathfrak{L}_{n,\epsilon,w}$ be the moduli space parametrizing:

(1) For all $v|p$ a basis $\rho_v : \mathcal{O}_{\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'}} \to \omega_{\mathcal{G}_v^{mod}}/\mathrm{Fil}\omega_{\mathcal{G}_v^{mod}}$ such that $\rho_v(1) = \mathrm{HT}(e_{v,4}) \mod p^{w'_v}$,
(2) For all $v \in I^c$, a basis $\nu_v : \mathcal{O}_{\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'}} \to \mathrm{Fil}\omega_{\mathcal{G}_v^{mod}}$ such that $\nu_v(1) = \mathrm{HT}(e_{v,3}) \mod p^{w'_v}$.

6.1.7. *Some groups.* The group $\prod_{v|p} \mathrm{GSp}_4(\mathcal{O}_{F_v}/p^{n_v})$ acts on $\mathfrak{X}_{K(p^n)}$ and $\mathfrak{X}_{K(p^n)}^{mod}$. The parabolic subgroup $\prod_{v\in I^c} \mathrm{B}(\mathcal{O}_{F_v}/p^{n_v}) \prod_{v\in I} \mathrm{Kli}(\mathcal{O}_{F_v}/p^{n_v})$ acts on $\mathfrak{X}_{K(p^n)}(\epsilon)$.

Let us denote by $\mathfrak{X}_{K(I,p^n)}(\epsilon)$ the quotient of $\mathfrak{X}_{K(p^n)}(\epsilon)$ by the action of this finite group. This is an admissible formal scheme.

We have maps $\mathrm{B}(\mathcal{O}_{F_v}/p^{n_v}) \to ((\mathcal{O}_{F_v}/p^{n_v})^\times)^2$ provided by the last two diagonal entries and $\mathrm{Kli}(\mathcal{O}_{F_v}/p^{n_v}) \to (\mathcal{O}_{F_v}/p^{n_v})^\times$ provided by the last diagonal entry.

We denote by $\mathfrak{T}^0_{w'}$ the formal group defined by

$$\mathfrak{T}^0_{w'}(R) = \prod_{v\in I}(1 + p^{w'_v}R) \prod_{v\in I^c}(1 + p^{w'_v}R)^2$$

for any admissible $\mathcal{O}_{\mathbf{C}_p}$-algebra R.

We denote by $\mathfrak{T}_{w'}$ the group

$$\mathfrak{T}_{w'}(R) = \prod_{v\in I} \mathcal{O}_{F_v}^\times(1 + p^{w'_v}R) \prod_{v\in I^c}(\mathcal{O}_{F_v}^\times(1 + p^{w'_v}R))^2$$

for any admissible $\mathcal{O}_{\mathbf{C}_p}$-algebra R.

Finally, we denote by $\mathfrak{T}_{n,w'}$ the fibre product

$$\mathfrak{T}_{w'} \times_{\mathfrak{T}_{w'}/\mathfrak{T}^0_{w'}} \prod_{v\in I^c} \mathrm{B}(\mathcal{O}_{F_v}/p^{n_v}) \prod_{v\in I} \mathrm{Kli}(\mathcal{O}_{F_v}/p^{n_v}).$$

6.1.8. *Torsors.* The map $\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'} \to \mathfrak{F}\mathfrak{L}_{n,\epsilon,w}$ is a $\mathfrak{T}^0_{w'}$-torsor. The group $\mathfrak{T}^0_{w'}$ acts on ρ_v and ν_v. This action extends to an action of $\mathfrak{T}_{n,w'}$ on $\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'} \to \mathfrak{X}_{K(p^n)}(\epsilon)$, compatible with the action of $\prod_{v\in I^c} \mathrm{B}(\mathcal{O}_{F_v}/p^{n_v}) \prod_{v\in I} \mathrm{Kli}(\mathcal{O}_{F_v}/p^{n_v})$ on $\mathfrak{X}_{K(p^n)}(\epsilon)$.

6.1.9. *Formal Banach sheaves.* Let A be a normal admissible $\mathcal{O}_{\mathbf{C}_p}$-algebra. Let $\kappa_A : \prod_{v\in I} \mathcal{O}^\times_{F_v} \prod_{v\in I^c} (\mathcal{O}^\times_{F_v})^2 \to A^\times$ be a character, which we assume is w'-analytic, in the sense that it extends to a pairing $\mathfrak{T}_{w'} \times \mathrm{Spf}\ A \to \mathbf{G}_m$.

We denote by $\pi_1 : \mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'} \to \mathfrak{F}\mathfrak{L}_{n,\epsilon,w}$ the projection. We can define an invertible sheaf of $\mathcal{O}_{\mathfrak{F}\mathfrak{L}_{n,\epsilon,w}}\hat{\otimes}A$-modules,

$$\mathfrak{L}^{\kappa_A} = ((\pi_1)_*\mathcal{O}_{\mathfrak{F}\mathfrak{L}^+_{n,\epsilon,w,w'}}\hat{\otimes}A)^{\mathfrak{T}^0_{w'}}$$

where the invariants are taken for the diagonal action.

We let $\pi_2 : \mathfrak{F}\mathfrak{L}_{n,\epsilon,w} \to \mathfrak{X}_{K(p^n)}(\epsilon)$. This is an affine map. We define a formal Banach sheaf $\mathfrak{G}^{\kappa_A,w} = (\pi_2)_*\mathfrak{L}^{\kappa_A}$ over $\mathfrak{X}_{K(p^n)}(\epsilon)$; this is independent of the choice of w', as is easily seen from the construction.

Finally, we let $\pi_3 : \mathfrak{X}_{K(p^n)}(\epsilon) \to \mathfrak{X}_{K(I,p^n)}(\epsilon)$, and we define $\mathfrak{F}^{\kappa_A,w} = ((\pi_3)_*\mathfrak{G}^{\kappa_A,w})^{\mathfrak{T}_{n,w}}$. This is a formal Banach sheaf over $\mathfrak{X}_{K(I,p^n)}(\epsilon)$.

6.1.10. *Some properties.* For each $v \in I$ we choose an element $i_v \in \{2,3\}$. Let $\mathfrak{X}_{K(p^n)}(\epsilon,(i_v))$ be the open subset of $\mathfrak{X}_{K(p^n)}(\epsilon)$ where $\mathrm{Fil}^{\mathrm{can}}_v$ is generated by $\mathrm{HT}(e_{i_v})$ for all $v \in I$.

Lemma 6.1.11. *The quasi-coherent sheaf $\mathfrak{G}^{\kappa_A,w}/p^{\inf_v w_v}$ restricted to $\mathfrak{X}_{K(p^n)}(\epsilon,(i_v))$ is an inductive limit of coherent sheaves which are extensions of the sheaf $\mathcal{O}_{\mathfrak{X}_{K(p^n)}(\epsilon,(i_v))}/p^{\inf_v w_v}$.*

Proof. This can be proved in the same way as [AIP15, Lem. 8.1.6.2]. □

Lemma 6.1.12. *The quasi-coherent sheaf $\mathfrak{G}^{\kappa_A,w}/p$ is a flat sheaf of $\mathcal{O}_{\mathfrak{X}_{K(p^n)}(\epsilon)}/p$-modules.*

Proof. See [Pil20, Lem. 12.6.2.1]. □

6.2. **Vanishing theorem.**

6.2.1. *The minimal compactification.* The main result of this subsection is Theorem 6.2.6. As in the proof of Theorem 4.2.1, we will use the minimal compactification, and in particular the facts that the pushforward of our sheaves to the minimal compactification are supported on open subsets that admit an explicit affine cover, and that the higher derived pushforwards from the toroidal to minimal compactifications of the cuspidal cohomology vanish.

We denote by $\mathfrak{X}^*_K$ the minimal compactification of $\mathfrak{Y}_K$. There is a natural map $\mathfrak{X}_K \to \mathfrak{X}^*_K$. The invertible sheaf $\det \omega_{\mathcal{G}}$ over $\mathfrak{X}_K$ descends to an invertible sheaf still denoted by $\det \omega_{\mathcal{G}}$ over $\mathfrak{X}^*_K$. Let $n = (n_v) \in \mathbf{Z}^{S_p}_{\geq 0}$. In this subsection we consider only the case that n_v is independent of v, and accordingly we will write n for n_v. We let $\mathfrak{X}^*_{K(p^n)}$ be the Stein factorization of the morphism: $\mathfrak{X}_{K(p^n)} \to \mathfrak{X}^*_K$. This

is a normal admissible formal scheme. In [PS16a, Cor. 1.4] it is proved that the determinant of the Hodge–Tate map on $\mathfrak{X}_{K(p^n)}$:

$$\Lambda^2\mathrm{HT} : \bigotimes_{v|p} \Lambda^2(\mathcal{O}_{F_v}/p^n\mathcal{O}_{F_v})^4 \to \bigotimes_{v|p} \det \omega_{\mathcal{G}_v}/p^n$$

is the pull back of a map denoted the same way:

$$\Lambda^2\mathrm{HT} : \bigotimes_{v|p} \Lambda^2(\mathcal{O}_{F_v}/p^n\mathcal{O}_{F_v})^4 \to \bigotimes_{v|p} \det \omega_{\mathcal{G}_v}/p^n$$

which is defined over $\mathfrak{X}^*_{K(p^n)}$.

Remark 6.2.2. Literally, the determinant of the Hodge–Tate map is a map:

$$\Lambda^{2[F:\mathbf{Q}]}(\mathcal{O}_F/p^n\mathcal{O}_F)^4 \to \det \omega_{\mathcal{G}}/p^n.$$

But using the action of $\mathcal{O}_F$, it is easy to see that it factors through the direct factor $\bigotimes_{v|p} \Lambda^2(\mathcal{O}_{F_v}/p^n\mathcal{O}_{F_v})^4$.

By [PS16a, §1.4] (for $F = \mathbf{Q}$, and the same construction for general F), there is a normal admissible formal scheme $\mathfrak{X}^{*-mod}_{K(p^n)} \to \mathfrak{X}^*_{K(p^n)}$ which is the normalization of a blow up and carries a locally free modification $\det \omega^{mod}_{\mathcal{G}} \subset \det \omega_{\mathcal{G}}$ such that:

(1) $p^{\frac{2[F:\mathbf{Q}]}{p-1}} \det \omega_{\mathcal{G}} \subset \det \omega^{mod}_{\mathcal{G}} \subset \det \omega_{\mathcal{G}}$.
(2) The Hodge–Tate map factorizes into a surjective map:

$$\bigotimes_{v|p} \Lambda^2(\mathcal{O}_{F_v}/p^n\mathcal{O}_{F_v})^4 \otimes \mathcal{O}_{\mathfrak{X}^{*-mod}_{K(p^n)}} \to \det \omega^{mod}_{\mathcal{G}}/p^{n-\frac{2[F:\mathbf{Q}]}{p-1}}.$$

The construction $\mathfrak{X}^{*-mod}_{K(p^n)}$ follows a similar procedure as the construction of $\mathfrak{X}^{mod}_{K(p^n)}$ explained in section 6.1.4: one can lift locally the map $\Lambda^2\mathrm{HT}$ to a map $\bigotimes_{v|p} \Lambda^2(\mathcal{O}^4_{F_v}) \otimes \mathcal{O}_{\mathfrak{X}^*_{K(p^n)}} \to \bigotimes_{v|p} \det \omega_{\mathcal{G}_v}$ and consider the normalization of the blow up of the ideal which is locally the product at all places v of the ideals generated by the coefficients of the above map at the place v. By the universal properties of blow-ups and normalizations, there is a map $\mathfrak{X}^{mod}_{K(p^n)} \to \mathfrak{X}^{*-mod}_{K(p^n)}$. Let $\epsilon = (\epsilon_v) \in \prod_{v|p}([0, n - \frac{2[F:\mathbf{Q}]}{p-1}] \cap \mathbf{Q})$. We denote by $\mathfrak{X}^*_{K(p^n)}(\epsilon) \to \mathfrak{X}^{*-mod}_{K(p^n)}$ the formal scheme defined by the condition:

- $\mathrm{HT}(e_{v,1}) \wedge \mathrm{HT}(e_{v,i}) \otimes_{v'|p, v' \neq v} \mathrm{HT}(e_{v',j_{v'}}) \wedge \mathrm{HT}(e_{v',k_{v'}}) \in p^{\epsilon_v} \det \omega^{mod}_{\mathcal{G}}/p^{n-\frac{2[F:\mathbf{Q}]}{p-1}}$ for all $v|p$ and $1 \leq i, j_{v'}, k_{v'} \leq 4$,
- $\mathrm{HT}(e_{v,2}) \wedge \mathrm{HT}(e_{v,i}) \otimes_{v'|p, v' \neq v} \mathrm{HT}(e_{v',j_{v'}}) \wedge \mathrm{HT}(e_{v',k_{v'}}) \in p^{\epsilon_v} \det \omega^{mod}_{\mathcal{G}}/p^{n-\frac{2[F:\mathbf{Q}]}{p-1}}$ for all $v \in I^c$ and $1 \leq i, j_{v'}, k_{v'} \leq 4$.

There is a Cartesian diagram (see the proof of [Pil20, Lem. 12.9.1.1]):

$$\begin{tikzcd}
\mathfrak{X}_{K(p^n)}(\epsilon) \arrow[r] \arrow[d] & \mathfrak{X}^{mod}_{K(p^n)} \arrow[d] \\
\mathfrak{X}^*_{K(p^n)}(\epsilon) \arrow[r] & \mathfrak{X}^{*-mod}_{K(p^n)}
\end{tikzcd}$$

By the proof of [Sch15, Thm. 4.3.1, pp 1029-30] (see also [PS16a, Thm. 1.16]), there is an integer N such that for all $n \geq N$, there is a normal admissible formal scheme

$\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}$ and a projective map $\mathfrak{X}^{*-mod}_{K(p^n)} \to \mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}$ which is an isomorphism on the associated analytic spaces and satisfies:

(1) The invertible sheaf $\det \omega^{mod}_{\mathcal{G}}$ descends to an ample invertible sheaf on $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}$.
(2) For all rational numbers $\epsilon > 0$, there is $n(\epsilon) \geq N$ such that if $n \geq n(\epsilon)$, then there are sections $s_{(i_v,j_v)_{v|p}} \in H^0(\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}, \det \omega^{mod}_{\mathcal{G}})$ satisfying $s_{(i_v,j_v)} = \otimes_{v|p}\mathrm{HT}(e_{v,i_v}) \wedge \mathrm{HT}(e_{v,j_v})$ in $\det \omega^{mod}_{\mathcal{G}}/p^{\epsilon}$ for all $1 \leq i_v, j_v \leq 4$.

Let $\epsilon = (\epsilon_v) \in (\mathbf{Q}_{>0})^{S_p}$. Let $n \geq \sup_v n(\epsilon_v)$. We define a formal scheme $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon) \to \mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}$ by the condition:

- for all $v|p$, for all $(i_{v'}, j_{v'})_{v'|p} \in (\{1,2,3,4\} \times \{1,2,3,4\})^{S_p}$, such that $i_v = 1$, we have $s_{(i_{v'},j_{v'})} \in p^{\epsilon_v} \det \omega^{mod}_{\mathcal{G}}$,
- for all $v \in I^c$, for all $(i_{v'}, j_{v'})_{v'|p} \in (\{1,2,3,4\} \times \{1,2,3,4\})^{S_p}$, such that $i_v = 2$, we have $s_{(i_{v'},j_{v'})} \in p^{\epsilon_v} \det \omega^{mod}_{\mathcal{G}}$.

We have a Cartesian diagram

$$\begin{tikzcd} \mathfrak{X}^{*}_{K(p^n)}(\epsilon) \arrow[r] \arrow[d] & \mathfrak{X}^{*-mod}_{K(p^n)} \arrow[d] \\ \mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon) \arrow[r] & \mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)} \end{tikzcd}$$

where both vertical maps are projective maps and induce isomorphisms on the associated analytic generic fibres.

For all $v \in I$, let $i_v \in \{2,3\}$. We define an open subspace $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon, (i_v))$ of $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon)$ by the condition that $s_{(i_v,4)_{v\in I},(3,4)_{v\in I^c}} \neq 0$.

We similarly define an open subspace $\mathfrak{X}^{*}_{K(p^n)}(\epsilon, (i_v))$ of $\mathfrak{X}^{*}_{K(p^n)}(\epsilon)$ by the condition that $\otimes_{v\in I}\mathrm{HT}(e_{v,i_v}) \wedge \mathrm{HT}(e_{v,4}) \otimes_{v\in I^c} \mathrm{HT}(e_{v,3}) \wedge \mathrm{HT}(e_{v,4}) \neq 0$ for all $v \in I$.

Lemma 6.2.3. *We have a projective map $\mathfrak{X}^{*}_{K(p^n)}(\epsilon, (i_v)) \to \mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon, (i_v))$ which is an isomorphism on the associated analytic adic spaces. Moreover, $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon, (i_v))$ is an affine formal scheme.*

Proof. The first point is clear. The second point follows from the ampleness of $\det \omega^{mod}_{\mathcal{G}}$ on $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon)$ and the fact that $\mathfrak{X}^{*-\mathrm{HT}}_{K(p^n)}(\epsilon, (i_v))$ is the open subscheme defined by the non-vanishing of a section of an ample sheaf. □

6.2.4. *Vanishing.* We have a map $\pi : \mathfrak{X}_{K(p^n)}(\epsilon) \to \mathfrak{X}^{*}_{K(p^n)}(\epsilon)$. We denote as usual by D the boundary divisor.

Proposition 6.2.5. *We have $\mathrm{R}^i\pi_*\mathcal{O}_{\mathfrak{X}_{K(p^n)}(\epsilon)}(-D) = 0$ for all $i > 0$.*

Proof. This can be proved in exactly the same way as [Pil20, Prop. 12.9.2.1] (which is the case $F = \mathbf{Q}$). □

Theorem 6.2.6. *Let $\epsilon = (\epsilon_v) \in \mathbf{Q}^{S_p}_{>0}$. Let $n = (n_v)$ with $\inf_v n_v \geq \sup_v n(\epsilon_v)$. The complex*

$$\mathrm{R}\Gamma(\mathfrak{X}_{K(p^n)}(\epsilon), \mathfrak{G}^{\kappa_A,w} \otimes (\det \omega^{mod}_{\mathcal{G}})^2(-D))[1/p]$$

has cohomology concentrated in degrees $[0, \#I]$.

Proof. Consider the hypercube $[2,3]^I$. We can associate to it a category denoted by $\mathcal{C}$. Its objects are the faces σ of the hypercube. By definition, a face is a product $\prod_{v\in I}\lambda_v$ where for each $v\in I$, $\lambda_v\in\{2,3,[2,3]\}$ (our convention is that faces are closed). There is a map $\sigma'\to\sigma$ between faces if σ' is included in σ.

We now define a functor $\mathcal{C}^{op}\to Op(\mathfrak{X}^{*-mod}_{K(p^n)}(\epsilon))$, where the target is the category of open subsets of $\mathfrak{X}^{*-mod}_{K(p^n)}(\epsilon)$ (whose morphisms are open immersions). It sends a face σ to U_σ, the intersection of all the formal schemes $\mathfrak{X}^{*-mod}_{K(p^n)}(\epsilon,(i_v))$ for $(i_v)_{v\in I}\in\sigma$ (we recall that $i_v\in\{2,3\}$).

Write $f:\mathfrak{X}_{K(p^n)}(\epsilon)\to\mathfrak{X}^{*-mod}_{K(p^n)}(\epsilon)$ for the map defined above. For all $\sigma\in\mathcal{C}$, consider the following Cartesian diagram:

$$\begin{tikzcd}
\mathfrak{X}_{K(p^n)}(\epsilon)_\sigma \arrow[r, "i"] \arrow[d, "f_\sigma"] & \mathfrak{X}_{K(p^n)}(\epsilon) \arrow[d, "f"] \\
U_\sigma \arrow[r, "j"] & \mathfrak{X}^{*-mod}_{K(p^n)}(\epsilon)
\end{tikzcd}$$

where j is the natural open immersion.

Let $Sh(\mathfrak{X}_{K(p^n)}(\epsilon))$ be the category of sheaves on $\mathfrak{X}_{K(p^n)}(\epsilon)$. We define a functor $\mathcal{C}\to Sh(\mathfrak{X}_{K(p^n)}(\epsilon))$ which sends σ to the sheaf

$$\mathfrak{G}_\sigma = i_*i^*(\mathfrak{G}^{\kappa_A,w}\otimes(\det\omega^{mod}_{\mathcal{G}})^2(-D)[1/p]).$$

We deduce from Lemmas 6.1.11 and 6.1.12, together with Proposition 6.2.5, that the sheaf $f_*\mathfrak{G}_\sigma$ is a small formal Banach sheaf and that $\mathrm{R}^if_*\mathfrak{G}_\sigma=0$ for all $i>0$. It follows from Theorem 6.1.3 (which applies because of Lemma 6.2.3) that $f_*\mathfrak{G}_\sigma$ is acyclic.

We deduce that the cohomology $\mathrm{R}\Gamma(\mathfrak{X}_{K(p^n)}(\epsilon),\mathfrak{G}^{\kappa_A,w}\otimes(\det\omega^{mod}_{\mathcal{G}})^2(-D)[1/p])$ is represented by the complex $C^\bullet$ concentrated in degree 0 to $\#I$, whose ith term is $\oplus_{\sigma,\dim\sigma=i}H^0(\mathfrak{X}_{K(p^n)}(\epsilon),\mathfrak{G}_\sigma)$ and whose differentials are alternating sums of the restriction maps $H^0(\mathfrak{X}_{K(p^n)}(\epsilon),\mathfrak{G}_{\sigma'})\to H^0(\mathfrak{X}_{K(p^n)}(\epsilon),\mathfrak{G}_\sigma)$ for $\sigma'\subset\sigma$, $\dim\sigma'=\dim\sigma-1$. □

6.3. **Sheaves of overconvergent and locally analytic modular forms: the analytic construction.** We now translate our previous formal constructions to the analytic setting, which is well adapted for the spectral theory.

6.3.1. *Analytic Hilbert–Siegel varieties.* This section is parallel to [Pil20, §12.7]. We let $\mathcal{X}_{K(p^n)}$ be the generic fibre of $\mathfrak{X}_{K(p^n)}$. We write $\mathcal{X}_K$ for the generic fibre of $\mathfrak{X}_K$. We let $\mathcal{X}_{K(p^n)}(\epsilon)\subset\mathcal{X}_{K(p^n)}$ be the generic fibre of $\mathfrak{X}_{K(p^n)}(\epsilon)$. We let $\mathcal{X}_{K(I,p^n)}(\epsilon)$ be the generic fibre of $\mathfrak{X}_{K(I,p^n)}(\epsilon)$. We now give a modular interpretation of this last space. Let A be the universal semi-abelian scheme and $\mathcal{G}$ be its p-divisible group. Let $\omega_{\mathcal{G}}$ be the conormal sheaf of A at the origin and let $\omega^+_{\mathcal{G}}\subset\omega_{\mathcal{G}}$ be the subsheaf of integral differentials (we use the slight abuse of notation to write $\omega_{\mathcal{G}}$ instead of ω_A). These are sheaves over $\mathcal{X}_K$ on the analytic site.

We let $\omega^{mod,+}_{\mathcal{G}}$ be the subsheaf of $\omega^+_{\mathcal{G}}$ generated by the image of the Hodge–Tate map. This is an étale sheaf over $\mathcal{X}_K$.

The fibres of the map $\mathcal{X}_{K(I,p^n)}(\epsilon)\to\mathcal{X}_K$ parametrize:

- For all $v \in I$, a subgroup $H_{v,n_v} \subset \mathcal{G}_v[p^{n_v}]$ which is locally for the étale topology isomorphic to $\mathbf{Z}/p^{n_v}\mathbf{Z}$ and is locally for the étale topology generated by an element $e_{v,1}$ which satisfies $\mathrm{HT}(e_{v,1}) = 0$ in $\omega_{\mathcal{G}_v}^{mod,+}/p^{\epsilon_v}$.
- For all $v \in I^c$, totally isotropic subgroups $H_{v,n_v} \subset L_{v,n_v} \subset \mathcal{G}_v[p^{n_v}]$ such that H_{v,n_v} is locally for the étale topology isomorphic to $\mathbf{Z}/p^{n_v}\mathbf{Z}$, L_{v,n_v} is locally for the étale topology isomorphic to $(\mathbf{Z}/p^{n_v}\mathbf{Z})^2$, and is locally for the étale topology generated by elements $e_{v,1}$ and $e_{v,2}$ which satisfy $\mathrm{HT}(e_{v,1}) = \mathrm{HT}(e_{v,2}) = 0$ in $\omega_{\mathcal{G}_v}^{mod,+}/p^{\epsilon_v}$.

We can define for all $v|p$ an étale sheaf $\mathrm{Fil}_v^{\mathrm{can}} = \mathrm{Im}(\mathrm{HT} : H_{v,n_v}^{\perp} \otimes \mathcal{O}^+_{\mathcal{X}_{K(I,p^n)}(\epsilon)} \to \omega_{\mathcal{G}_v}^{mod,+}/p^{\epsilon_v})$. This is a étale locally free sheaf of $\mathcal{O}^+_{\mathcal{X}_{K(I,p^n)}(\epsilon)}/p^{\epsilon_v}$-modules of rank 1. We let $\mathrm{Gr}_v^{\mathrm{can}} = \omega_{\mathcal{G}_v}^{mod,+}/(p^{\epsilon_v} + \mathrm{Fil}_v^{\mathrm{can}})$.

We have isomorphisms deduced from the Hodge–Tate map:

- $\mathrm{HT} : H_{v,n_v}^D \otimes \mathcal{O}^+_{\mathcal{X}_{K(I,p^n)}(\epsilon)}/p^{\epsilon_v} \to \mathrm{Gr}_v^{\mathrm{can}}$ for all $v|p$,
- $\mathrm{HT} : (L_{v,n_v}/H_{v,n_v})^D \otimes \mathcal{O}^+_{\mathcal{X}_{K(I,p^n)}(\epsilon)}/p^{\epsilon_v} \to \mathrm{Fil}_v^{\mathrm{can}}$ for all $v \in I^c$.

We let $\mathcal{FL}_{K(I,p^n),\epsilon,w} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ be the moduli space of flags $\mathrm{Fil}\omega_{\mathcal{G}} \subset \omega_{\mathcal{G}}$ satisfying $\mathrm{Fil}\omega_{\mathcal{G}_v} \cap \omega_{\mathcal{G}_v}^{mod,+}/p^{w_v} = \mathrm{Fil}_v^{\mathrm{can}}/p^{w_v}$.

We let $\mathcal{X}_{K(I,p^n)^+}(\epsilon) \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ be the étale cover parametrizing trivializations:

- $\mathbf{Z}/p^{n_v}\mathbf{Z}e_{v,4} \xrightarrow{\sim} H_{v,n_v}^D$ for all $v|p$,
- $\mathbf{Z}/p^{n_v}\mathbf{Z}e_{v,3} \xrightarrow{\sim} (L_{v,n_v}/H_{v,n_v})^D$ for all $v \in I^c$.

We let $\mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'} \to \mathcal{FL}_{K(I,p^n),\epsilon,w} \times_{\mathcal{X}_{K(I,p^n)}(\epsilon)} \mathcal{X}_{K(I,p^n)^+}(\epsilon)$ be the moduli space of trivializations of:

- for all $v|p$, $\rho_v : \mathcal{O}_{\mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'}} \to \mathrm{Gr}\omega_{\mathcal{G}_v} = \omega_{\mathcal{G}_v}/\mathrm{Fil}\,\omega_{\mathcal{G}_v}$ such that $\rho_v(1) = \mathrm{HT}(e_{v,4})$ modulo $p^{w'_v}$.
- for all $v \in I^c$, $\nu_v : \mathcal{O}_{\mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'}} \to \mathrm{Fil}\omega_{\mathcal{G}_v}$ such that $\nu_v(1) = \mathrm{HT}(e_{v,3})$ modulo $p^{w'_v}$.

We can connect these definitions with the constructions of the previous sections. Let $\mathcal{FL}_{n,\epsilon,w} \to \mathcal{X}_{K(p^n)}(\epsilon)$ be the analytic space associated to $\mathfrak{FL}_{n,\epsilon,w}$. Let $\mathcal{FL}^+_{n,\epsilon,w,w'}$ be the analytic space associated to $\mathfrak{FL}^+_{n,\epsilon,w,w'}$.

Lemma 6.3.2. *We have*

$$\mathcal{FL}_{n,\epsilon,w} = \mathcal{FL}_{K(I,p^n),\epsilon,w} \times_{\mathcal{X}_{K(I,p^n)}(\epsilon)} \mathcal{X}_{K(p^n)}(\epsilon)$$

and

$$\mathcal{FL}^+_{n,\epsilon,w,w'} = \mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'} \times_{\mathcal{X}_{K(I,p^n)^+}(\epsilon)} \mathcal{X}_{K(p^n)}(\epsilon).$$

Proof. This follows from the definitions. □

6.3.3. *Banach sheaves.* We let $\mathcal{L}^{\kappa_A}$ be the invertible sheaf over $\mathcal{FL}_{n,\epsilon,w} \times \mathrm{Spa}(A[1/p], A)$ associated to $\mathfrak{L}^{\kappa_A}$. We let $\mathcal{G}^{\kappa_A,w}$ be the Banach sheaf over $\mathcal{X}_{K(p^n)}(\epsilon)$ associated to $\mathfrak{G}^{\kappa_A,w}$. We let $\mathcal{F}^{\kappa_A,w}$ be the Banach sheaf over $\mathcal{X}_{K(I,p^n)}(\epsilon)$ attached to $\mathfrak{F}^{\kappa_A,w}$. A direct definition of $\mathcal{F}^{\kappa_A,w}$ is the following. Let $\pi : \mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ be the affine projection. Let $\mathcal{T}_{w'}$ be the generic fibre of $\mathfrak{T}_{w'}$. This group acts naturally on $\mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'}$, trivially on $\mathcal{X}_{K(I,p^n)}(\epsilon)$, and the morphism π is equivariant for

the action. It follows from the definitions that

$$\mathcal{F}^{\kappa_A,w} = (\pi_* \mathcal{O}_{\mathcal{FL}^+_{K(I,p^n),\epsilon,w,w'}} \hat{\otimes} A)^{\mathcal{T}_{w'}}$$

where the invariants are for the diagonal action (with the action on the second factor being via κ_A).

6.3.4. *Locally analytic overconvergent cohomology.* We define the n,ϵ-convergent, cuspidal w-analytic cohomology of weight parametrized by A to be:

$$C_{cusp}(n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p}) := \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)).$$

For $\epsilon' \geq \epsilon$, $n' \geq n$, $w' \geq w$, we have maps: $C_{cusp}(n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p}) \to C_{cusp}(n',\epsilon',w',\kappa_A \otimes (2,2)_{v|p})$.

Passing to the limit over n,ϵ,w, we define the ith cohomology groups of cuspidal, overconvergent, locally analytic cohomology of weight parametrized by A:

$$H^i_{cusp}(\dagger,\kappa_A \otimes (2,2)_{v|p}) = \varinjlim H^i(C_{cusp}(n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p})).$$

6.3.5. *Properties of locally analytic overconvergent cohomology.*

Proposition 6.3.6. *The complex $C_{cusp}(n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p})$ is represented by a bounded complex of projective Banach $A[1/p]$-modules.*

Proof. This follows easily by considering a Čech complex; see [Pil20, Prop. 12.8.2.1]. □

Proposition 6.3.7. *The cohomology $H^i_{cusp}(\dagger,\kappa_A \otimes (2,2)_{v|p})$ vanishes for $i \notin [0,\#I]$.*

Proof. This follows from Theorem 6.2.6. □

6.3.8. *Descent.* We now assume that the character

$$\kappa_A : \prod_{v\in I} \mathcal{O}^\times_{F_v} \prod_{v\in I^c} (\mathcal{O}^\times_{F_v})^2 \to A^\times$$

is trivial on the torsion subgroup of $\prod_{v\in I} \mathcal{O}^\times_{F_v} \prod_{v\in I^c} (\mathcal{O}^\times_{F_v})^2$ (of order prime to p since $p > 2$).

The group $(\mathcal{O}_F)^{\times,+}_{(p)}$ acts on $\mathcal{X}_{K(I,p^n)}$, and the action factors through a finite group. We let $\mathcal{X}^{G_1}_{K(I,p^n)}$ be the quotient. The action of $(\mathcal{O}_F)^{\times,+}_{(p)}$ can be lifted to the sheaf $\mathcal{F}^{\kappa_A,w}$ by setting

$$x : x^*\mathcal{F}^{\kappa_A,w} \to \mathcal{F}^{\kappa_A,w}$$

for all $x \in (\mathcal{O}_F)^{\times,+}_{(p)}$, to be the composition of the tautological isomorphism (the polarization is not used in the construction of the sheaf) and multiplication by the character $d : (\mathcal{O}_F)^{\times,+}_{(p)} \to \Lambda^\times_I \to A^\times$ of §4.4.2.

We denote by $\mathcal{F}^{\kappa_A^{G_1},w}$ the descended sheaf on $\mathcal{X}^{G_1}_{K(I,p^n)}$. We let

$$C_{cusp}(G_1,n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p})$$

be the cohomology of the sheaf

$$\mathcal{F}^{\kappa_A^{G_1},w} \otimes (\det \omega_{\mathcal{G}})^2(-D)$$

over $\mathcal{X}^{G_1}_{K(I,p^n)}(\epsilon)$. This is a direct factor of $C_{cusp}(n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p})$. We also let $H^i_{cusp}(G_1,\dagger,\kappa_A \otimes (2,2)_{v|p}) = \varinjlim H^i(C_{cusp}(G_1,n,\epsilon,w,\kappa_A \otimes (2,2)_{v|p}))$.

6.3.9. *Spectral theory: construction of the operator* $U_{v,2}$. Firstly let $v \in I$. We define an analytic adic space $s_1 : \mathcal{C}_{v,2} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ which parametrizes isogenies $A \to A'$ with associated Barsotti–Tate group $\mathcal{G} \to \mathcal{G}'$ whose kernel is a group $M_v \subset \mathcal{G}_v[p^2]$ which:

- is totally isotropic and locally isomorphic to $(\mathcal{O}_{F_v}/p\mathcal{O}_{F_v})^2 \oplus \mathcal{O}_{F_v}/p^2\mathcal{O}_{F_v}$,
- has trivial intersection with H_{v,n_v}.

There is a second projection $s_2 : \mathcal{C}_{v,2} \to \mathcal{X}_{K(I,p^{n'})}(\epsilon')$ where:

- $n' = (n'_{v'})_{v'|p}$ where $n'_v = n_v + 1$, and $n'_{v'} = n_{v'}$ if $v' \neq v$.
- $\epsilon' = (\epsilon_{v'})_{v'|p}$ where $\epsilon'_v = \epsilon_v + 1$, and $\epsilon'_{v'} = \epsilon_{v'}$ if $v' \neq v$.

This map is provided by sending (A, A') to A', equipped with the subgroups:

- $H'_{v',n_{v'}} = \mathrm{Im}(H_{v',n_{v'}})$ for all $v' \neq v$,
- $L'_{v',n_{v'}} = \mathrm{Im}(L_{v',n_{v'}})$ for all $v' \in I^c$,
- $H'_{v,n_v+1} = \mathrm{Im}(p^{-1}H_{v,n_v})$ where $p^{-1}H_{v,n_v}$ is the pre-image in $\mathcal{G}_v[p^{n_v+1}]$ of H_{v,n_v}.

One checks as in [Pil20, Lem. 13.2.1.1] (see also Lemma 6.3.13 below) that the image of s_2 lands in $\mathcal{X}_{K(I,p^{n'})}(\epsilon')$. The natural map $\omega_{\mathcal{G}'} \to \omega_{\mathcal{G}}$ induces a natural map $s_2^*\mathcal{F}^{\kappa_A,w'} \to s_1^*\mathcal{F}^{\kappa_A,w}$, where $w' = (w_{v'})$ with $w'_{v'} = w_{v'}$ if $v' \neq v$, and $w'_v = w_v + 1$. (See [Pil20, Lem. 13.2.2.1].) We deduce that there is a normalized map $s_2^*\mathcal{F}^{\kappa_A,w'} \otimes (\det \omega_{\mathcal{G}})^2(-D) \to s_1^*\mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)$ obtained by taking the tensor product of the above map and the normalized map (by p^{-2}) $s_2^*(\det \omega_{\mathcal{G}})^2(-D) \to s_1^*(\det \omega_{\mathcal{G}})^2(-D)$.

We can therefore construct a Hecke operator $U_{v,2} : \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}, \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$ by the following composition:

$$\mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^{n'})}(\epsilon'), \mathcal{F}^{\kappa_A,w'} \otimes (\det \omega_{\mathcal{G}})^2(-D))$$

$$\to \mathrm{R}\Gamma(\mathcal{C}_{v,2}, s_2^*\mathcal{F}^{\kappa_A,w'} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to$$

$$\mathrm{R}\Gamma(\mathcal{C}_v, s_1^*\mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), (s_1)_* s_1^*\mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$$

$$\xrightarrow{p^{-3}\mathrm{Tr}} \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)).$$

Now let $v \in I^c$. We define an analytic adic space $s_1 : \mathcal{C}_{v,2} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ which parametrizes isogenies $\mathcal{G} \to \mathcal{G}'$ whose kernel is a group $M_v \subset \mathcal{G}_v[p^2]$ which:

- is totally isotropic and locally isomorphic to $(\mathcal{O}_{F_v}/p\mathcal{O}_{F_v})^2 \oplus \mathcal{O}_{F_v}/p^2\mathcal{O}_{F_v}$, and
- locally in the étale topology there is a symplectic isomorphism $(\mathcal{G}_v)[p^\infty] \simeq (F_v/\mathcal{O}_{F_v})^4$ such that
  - M_v is generated by $p^{-1}e_{v,2}, p^{-1}e_{v,3}, p^{-2}e_{v,4}$,
  - H_{v,n_v} is generated by $p^{-n_v}e_{v,1}$, and
  - L_{v,n_v} is generated by $p^{-n_v}e_{v,1}$ and $p^{-n_v}e_{v,2}$.

There is a second projection $s_2 : \mathcal{C}_{v,2} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$, sending $(\mathcal{G}, \mathcal{G}')$ to $\mathcal{G}'$, equipped with the subgroups:

- $H'_{v',n_{v'}} = \mathrm{Im}(H_{v',n_{v'}})$ for all $v' \neq v$.
- $L'_{v',n_{v'}} = \mathrm{Im}(L_{v',n_{v'}})$ for all $v' \in I^c$, $v' \neq v$.

- In the notation above, H'_{v,n_v} is the group generated by the image in $\mathcal{G}'$ of $p^{-n_v}e_{v,1}$ and L'_{v,n_v} is the group generated by the image of $p^{-n_v}e_{v,1}$ and $p^{-n_v-1}e_{v,2}$. One checks easily that these groups only depend on M_v, H_{v,n_v} and L_{v,n_v} (and not on the choice of symplectic basis).

Again, there is a natural map $s_2^*\mathcal{F}^{\kappa_A,w} \to s_1^*\mathcal{F}^{\kappa_A,w}$. (See [Pil20, Lem. 13.2.2.1] and [AIP15, §6.2].) We deduce that there is a normalized map $s_2^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D) \to s_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)$ obtained by taking the tensor product of the above map and the normalized map (by p^{-2}) $s_2^*(\det\omega_{\mathcal{G}})^2(-D) \to s_1^*(\det\omega_{\mathcal{G}})^2(-D)$.

We can therefore construct a Hecke operator $U_{v,2} : \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}, \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D))$ by the following composition:

$$\mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{C}_{v,2}, s_2^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to$$

$$\mathrm{R}\Gamma(\mathcal{C}_{v,2}, s_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), (s_1)_*s_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D))$$

$$\xrightarrow{p^{-3}\mathrm{Tr}} \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)).$$

Remark 6.3.10. When $v \in I^c$ we observe that $U_{v,2}$ itself is not improving analyticity and convergence in the v direction (while it visibly does so in the case $v \in I$). We next define an operator $U_{v,1}$ when $v \in I^c$. We will then show that the composite operator $U_{v,1}U_{v,2}$ improves analyticity and convergence. (This is related to our needing to use both the operators T_v and $T_{v,1}$ at places $v \in I^c$ in §4.)

6.3.11. *Spectral theory: construction of the operator* $U_{v,1}$. We let $v \in I^c$. We define an analytic adic space $t_1 : \mathcal{C}_{v,1} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ which parametrizes isogenies $A \to A'$ with associated Barsotti–Tate groups $\mathcal{G} \to \mathcal{G}'$ whose kernel is a group $M_v \subset \mathcal{G}_v[p]$ which:

- is totally isotropic and locally isomorphic to $(\mathcal{O}_{F_v}/p\mathcal{O}_{F_v})^2$,
- has trivial intersection with L_{v,n_v}.

There is a second projection $t_2 : \mathcal{C}_{v,1} \to \mathcal{X}_{K(I,p^n)}(\epsilon)$, given by sending (A, A') to A', equipped with the subgroups:

- $H'_{v',n_{v'}} = \mathrm{Im}(H_{v',n_{v'}})$ for all v',
- $L'_{v',n_{v'}} = \mathrm{Im}(L_{v',n_{v'}})$ for all $v' \in I^c$.

There is a natural map $t_2^*\mathcal{F}^{\kappa_A,w} \to t_1^*\mathcal{F}^{\kappa_A,w}$ (again see [Pil20, Lem. 13.2.2.1] and [AIP15, §6.2]). We deduce that there is a map $t_2^*\mathcal{F}^{\kappa_A,w'}\otimes(\det\omega_{\mathcal{G}})^2(-D) \to t_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)$ obtained by taking the tensor product of the above map and the map $t_2^*(\det\omega_{\mathcal{G}})^2(-D) \to t_1^*(\det\omega_{\mathcal{G}})^2(-D)$.

We can therefore construct a Hecke operator $U_{v,1} : \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}, \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D))$ by the following composition:

$$\mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{C}_{v,1}, t_2^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to$$

$$\mathrm{R}\Gamma(\mathcal{C}_{v,1}, t_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), (t_1)_*t_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D))$$

$$\xrightarrow{p^{-3}\mathrm{Tr}} \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det\omega_{\mathcal{G}})^2(-D)).$$

6.3.12. *Spectral theory: construction of the operator $U_{v,1}U_{v,2}$.* Let $v \in I^c$. We now consider the composite operator $U_{v,1}U_{v,2}$. Our main task it to show that this operator improves convergence and analyticity in the v-direction. We begin by giving the correspondence corresponding to this composite.

We define an analytic adic space $u_1 : \mathcal{C}_v \to \mathcal{X}_{K(I,p^n)}(\epsilon)$ which parametrizes isogenies $A \to A'$ with associated Barsotti–Tate groups $\mathcal{G} \to \mathcal{G}'$ whose kernel is a group $M_v \subset \mathcal{G}_v[p^3]$ which:

- is totally isotropic and locally isomorphic to $\mathcal{O}_{F_v}/p\mathcal{O}_{F_v} \oplus \mathcal{O}_{F_v}/p^2\mathcal{O}_{F_v} \oplus \mathcal{O}_{F_v}/p^3\mathcal{O}_{F_v}$,
- locally in the étale topology there is a symplectic isomorphism $(\mathcal{G}_v)[p^\infty] \simeq (F_v/\mathcal{O}_{F_v})^4$ such that
  - M_v is generated by $p^{-1}e_{v,2}, p^{-2}e_{v,3}, p^{-3}e_{v,4}$,
  - H_{v,n_v} is generated by $p^{-n_v}e_{v,1}$, and
  - L_{v,n_v} is generated by $p^{-n_v}e_{v,1}$ and $p^{-n_v}e_{v,2}$.

There is a second projection $u_2 : \mathcal{C}_v \to \mathcal{X}_{K(I,p^{n'})}(\epsilon')$, given by sending (A, A') to A', equipped with the subgroups:

- $H'_{v',n_{v'}} = \mathrm{Im}(H_{v',n_{v'}})$ for all $v' \neq v$,
- $L'_{v',n_{v'}} = \mathrm{Im}(L_{v',n_{v'}})$ for all $v' \in I^c$, $v' \neq v$,
- In the notation above, H'_{v,n_v+1} is the group generated by the image in $\mathcal{G}'$ of $p^{-n_v-1}e_{v,1}$ and L'_{v,n_v+1} is the group generated by the image of $p^{-n_v-1}e_{v,1}$ and $p^{-n_v-2}e_{v,2}$. One checks easily that these groups only depend on M_v, H_{v,n_v} and L_{v,n_v}.

Lemma 6.3.13. *The image of u_2 lands in $\mathcal{X}_{K(I,p^{n'})}(\epsilon')$.*

Proof. We argue in the same way as in the proof of [Pil20, Lem. 13.2.1.1]. We fix symplectic bases $(e_{v,i})_{1\leq i\leq 4}$ of $T_p(\mathcal{G})$, $(e'_{v,i})_{1\leq i\leq 4}$ of $T_p(\mathcal{G}')$, $(f_i)_{1\leq i\leq 2}$ of $\omega^{mod}_{\mathcal{G}_v}$, and $(f'_i)_{1\leq i\leq 2}$ of $\omega^{mod}_{\mathcal{G}'_v}$ (compatible with the canonical filtration) such that there is a commutative diagram:

$$\begin{tikzcd} T_p(\mathcal{G}_v) \arrow[r, "\mathrm{diag}(1,p,p^2,p^3)"] \arrow[d, "\mathrm{HT}"] & T_p(\mathcal{G}'_v) \arrow[d, "\mathrm{HT}"] \\ \omega^{mod}_{\mathcal{G}_v} \arrow[r, "\mathrm{diag}(p^2,p^3)"] & \omega^{mod}_{\mathcal{G}'} \end{tikzcd}$$

By definition we have that $\mathrm{HT}(e_{v,1}), \mathrm{HT}(e_{v,2}) \in p^{\epsilon_v}\omega^{mod}_{\mathcal{G}_v}$. On the other hand, $\mathrm{HT}(e_{v,3}), \mathrm{HT}(e_{v,4})$ generate $\omega^{mod}_{\mathcal{G}_v}$ and $\mathrm{HT}(e'_{v,3}), \mathrm{HT}(e'_{v,4})$ generate $\omega^{mod}_{\mathcal{G}'_v}$. The group L_{v,n_v+1} is generated by $\mathrm{diag}(1,p,p^2,p^3)\cdot e_{v,1} = e'_{v,1}$ and $\mathrm{diag}(1,p,p^2,p^3)\cdot p^{-1}e_{v,2} = e'_{v,2}$. Therefore we deduce that $\mathrm{HT}(e'_{v,1}), \mathrm{HT}(e'_{v,2}) \in p^{\epsilon_v+1}\omega^{mod}_{\mathcal{G}}$. □

There is again a natural map $u_2^*\mathcal{F}^{\kappa_A,w'} \to u_1^*\mathcal{F}^{\kappa_A,w}$ where $w' = (w_{v'})$ with $w'_{v'} = w_{v'}$ is $v' \neq v$ and $w'_v = w_v + 1$; see [Pil20, Lem. 13.2.2.1] and [AIP15, §6.2]. We deduce that there is a normalized map $u_2^*\mathcal{F}^{\kappa_A,w'} \otimes (\det\omega_{\mathcal{G}})^2(-D) \to u_1^*\mathcal{F}^{\kappa_A,w} \otimes (\det\omega_{\mathcal{G}})^2(-D)$ obtained by taking the tensor product of the above map and the normalized map (by p^{-2}) $u_2^*(\det\omega_{\mathcal{G}})^2(-D) \to u_1^*(\det\omega_{\mathcal{G}})^2(-D)$.

We can therefore construct a Hecke operator $U_{v,2}U_{v,1} : \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}, \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$ by the following composition:

$$\mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w}\otimes(\det \omega_{\mathcal{G}})^2(-D)) \overset{res}{\to} \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^{n'})}(\epsilon'), \mathcal{F}^{\kappa_A,w'}\otimes(\det \omega_{\mathcal{G}})^2(-D))$$

$$\to \mathrm{R}\Gamma(\mathcal{C}_v, u_2^*\mathcal{F}^{\kappa_A,w'} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to$$

$$\mathrm{R}\Gamma(\mathcal{C}_v, u_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), (u_1)_*u_1^*\mathcal{F}^{\kappa_A,w}\otimes(\det \omega_{\mathcal{G}})^2(-D))$$

$$\overset{p^{-6}\mathrm{Tr}}{\to} \mathrm{R}\Gamma(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$$

We now set $U^I = \prod_{v\in I} U_{v,2} \prod_{v\in I^c} U_{v,1}U_{v,2}$.

Lemma 6.3.14. *The operator U^I acting on $C_{cusp}(G_1, n, \epsilon, w, \kappa_A \otimes (2,2)_{v|p})$ is compact. Moreover, for $n+1 = (n_v+1)_{v|p}$, $\epsilon+1 = (\epsilon_v+1)_v$ and $w+1 = (w_v+1)_v$ we have a factorization (where the vertical maps are the natural restriction maps):*

$$\begin{tikzcd}
C_{cusp}(G_1, n, \epsilon, w, \kappa_A \otimes (2,2)_{v|p}) \arrow[r, "U^I"] \arrow[d] & C_{cusp}(G_1, n, \epsilon, w, \kappa_A \otimes (2,2)_{v|p}) \arrow[d] \\
C_{cusp}(G_1, n+1, \epsilon+1, w+1, \kappa_A \otimes (2,2)_{v|p}) \arrow[r, "U^I"] \arrow[ur] & C_{cusp}(G_1, n+1, \epsilon+1, w+1, \kappa_A \otimes (2,2)_{v|p})
\end{tikzcd}$$

Proof. By construction, the action of U^I can be factored into

$$\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A^{G_1},w}\otimes(\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p^{n+1})}(\epsilon+1), \mathcal{F}^{\kappa_A^{G_1},w+1}\otimes(\det \omega_{\mathcal{G}})^2(-D))$$

$$\overset{\tilde{U}^I}{\to} \mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A^{G_1},w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$$

where $n+1 = (n_v+1)_{v|p}$ and $\epsilon+1 = (\epsilon_v+1)_v$, $w+1 = (w_v+1)_v$. It is enough to show that the map $\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa_A^{G_1},w} \otimes (\det \omega_{\mathcal{G}})^2(-D)) \to \mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p^{n+1})}(\epsilon+1), \mathcal{F}^{\kappa_A^{G_1},w+1} \otimes (\det \omega_{\mathcal{G}})^2(-D))$ is compact. This follows by consideration of an appropriate Čech complex, as in [Pil20, Lem. 13.2.4.1]. □

6.3.15. *Spectral theory: local constancy of the Euler Characteristics.* Let W_{cl} be the set of weights $\kappa = ((k_v, l_v))_{v|p} \in \mathbf{Z}^{S_p}$, with $l_v = 2$ if $v \in I$, $k_v \equiv l_v \equiv 2 \bmod (p-1)$ for all $v|p$. It is equipped with the p-adic topology.

For all $\kappa \in W_{cl}$, we let $C_{cusp}(G_1, n, \epsilon, w, \kappa)$ be n, ϵ-convergent, w-analytic cohomology of weight κ and we set $H^i_{cusp}(G_1, \dagger, \kappa) = \varinjlim H^i(C_{cusp}(G_1, n, \epsilon, w, \kappa))$. In other words, following the notation of §6.3.8, we have $A = \mathbf{C}_p$ and $\kappa_A = \kappa \otimes (-2,-2)_{v|p}$. It follows from Lemma 6.3.14 that the cohomology groups $e(U^I)H^i_{cusp}(G_1, \dagger, \kappa)$ are finite-dimensional. The following standard consequence of our constructions will be crucial in our comparison in §6.6 of the complexes constructed in §4 and the overconvergent cohomology we are considering in this section.

Theorem 6.3.16. *The map*

$$\begin{array}{rcl} W_{cl} & \to & \mathbf{Z} \\ \kappa & \mapsto & \sum_i (-1)^i \dim e(U^I)H^i_{cusp}(G_1, \dagger, \kappa) \end{array}$$

is locally constant.

Proof. This follows from Coleman's theory [Col97, §A5], as in [Pil20, §13.4]. Indeed, there is a perfect complex $C^\bullet$ interpolating $C_{cusp}(G_1, n, \epsilon, w, \kappa)$ over the spectral variety, and the dimensions of the slope zero parts of the C^i are locally constant. □

Remark 6.3.17. In particular if $\#I = 1$, we deduce that

$$\kappa \mapsto \dim e(U^I)H^0_{cusp}(G_1, \dagger, \kappa) - \dim e(U^I)H^1_{cusp}(G_1, \dagger, \kappa)$$

is locally constant. We will use this in §6.6 to reduce the comparison of ordinary and overconvergent cohomology to the case of high weight, where the control theorems proved in §4 apply.

6.4. **Locally analytic overconvergent classes and algebraic overconvergent classes.** Let $\kappa = ((k_v, l_v))_{v|p}$ with $l_v = 2$ if $v \in I$, $k_v \equiv l_v \equiv 2 \bmod (p-1)$ be a dominant algebraic weight.

Proposition 6.4.1. *On $H^i_{cusp}(G_1, \dagger, \kappa)$, the slopes of $(U_{v,1})_{v\in I^c}$ and $(U_{v,2})_{v|p}$ are ≥ -3. On $H^0_{cusp}(G_1, \dagger, \kappa)$ they are ≥ 0.*

Proof. The proof of [Pil20, Prop. 13.3.1.1] goes through essentially without change. □

Below, we denote by $\mathcal{F}^{\kappa,w^-} = \varinjlim_{w'<w} \mathcal{F}^{\kappa,w'}$.

Proposition 6.4.2. *Let $\kappa = ((k_v, l_v))_{v|p}$ with $l_v = 2$ if $v \in I$, $k_v = l_v = 2 \bmod (p-1)$ be a dominant algebraic weight. There is a relative analytic BGG resolution:*

$$0 \to \omega^\kappa(-D) \to \mathcal{F}^{\kappa,w^-} \otimes (\det \omega_{\mathcal{G}})^2(-D) \to$$
$$\bigoplus_{s\in W^{(1)}} \mathcal{F}^{s\bullet\kappa,w^-} \otimes (\det \omega_{\mathcal{G}})^2(-D) \to \ldots \to \bigoplus_{s\in W^{(d)}} \mathcal{F}^{s\bullet\kappa,w^-} \otimes (\det \omega_{\mathcal{G}})^2(-D) \to 0$$

where W is the Weyl group of $\mathrm{GL}_2(F \otimes_{\mathbf{Q}} \mathbf{Q}_p)$, $W^{(i)}$ stands for the elements of length i in W, and $\bullet$ is the twisted Weyl action.

Proof. This is a relative version of the main result of [Jon11], and is proved in [AIP15, §7.2]. (Note though that there is a minor error there; one needs to replace $\mathcal{F}^{\kappa,w}$ with $\mathcal{F}^{\kappa,w^-}$ as defined above, but having made this change, the arguments go through unchanged.) □

The actions of $U_{v,1}$ and $U_{v,2}$ by cohomological correspondences on the sheaf $\mathcal{F}^{\kappa,w^-} \otimes (\det \omega_{\mathcal{G}})^2(-D)$ restrict to actions on the subsheaf $\omega^\kappa(-D)$ (and the action of U^I is compact on the cohomology).

Corollary 6.4.3. *Let $\kappa = ((k_v, l_v))_{v|p}$ with $l_v = 2$ if $v \in I$, $k_v \equiv l_v \equiv 2 \bmod (p-1)$ be a dominant algebraic weight. Then the map*

$$e(U^I)H^i(\mathcal{X}_{K(I,p^n)}(\epsilon), \omega^\kappa(-D)) \to e(U^I)H^i(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{\kappa,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$$

is an isomorphism for $i = 0$ and injective if $i = 1$. It is an isomorphism for $i = 1$ if we further assume that $k_v - l_v \geq 3$ for all $v|p$.

Proof. Proposition 6.4.2 gives a spectral sequence $E_1^{p,q} = \oplus_{s\in W^{(p)}} H^q(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{s\bullet\kappa,w} \otimes (\det \omega_{\mathcal{G}})^2(-D))$ converging to $H^{p+q}(\mathcal{X}_{K(I,p^n)}(\epsilon), \omega^\kappa(-D))$. We shall see that the ordinary projector kills the terms $E_1^{p,q}$ of the spectral sequence for $p > 1$ under a suitable normalization of the action of the Hecke operators and suitable assumptions on the weight κ. We analyze the differentials of proposition 6.4.2:

$\bigoplus_{s\in W^{(i)}} \mathcal{F}^{s\bullet\kappa,w^-} \otimes (\det\omega_{\mathcal{G}})^2(-D) \to \bigoplus_{s\in W^{(i+1)}} \mathcal{F}^{s\bullet\kappa,w^-} \otimes (\det\omega_{\mathcal{G}})^2(-D)$. We let $W = \prod_{v|p}\{1_v, w_v\}$ with $\ell(w_v) = 1$. For any subset J of places diving p, we let $w_J = \prod_{v\in J} w_v$. The above map is given by the product of the maps $\theta_{s,s'} : \mathcal{F}^{s\bullet\kappa,w^-} \otimes (\det\omega_{\mathcal{G}})^2(-D) \to \mathcal{F}^{s'\bullet\kappa,w^-} \otimes (\det\omega_{\mathcal{G}})^2(-D)$ for $s = w_J$ (for a subset J of cardinality i) and $s' = w_{J\cup\{v\}}$ for $v \notin J$. By [AIP15, §7.3], this map induces on cohomology an equivariant map for the operators $U_{w,i}$ for $w \neq v$ and $U_{v,1}$; and on the other hand, we have $U_{v,2} \circ \theta_{s,s'} = p^{(k_v-l_v)+1}\theta_{s,s'} \circ U_{v,2}$. A way to interpret this relation is to say that the spectral sequence is equivariant for the action of Hecke operators, if the standard action of $U_{v,2}$ on $H^i(\mathcal{X}_{K(I,p^n)}(\epsilon), \mathcal{F}^{w_J\bullet\kappa,w} \otimes (\det\omega_{\mathcal{G}})^2(-D))$ is twisted by multiplication by $p^{k_v-l_v+1}$ if $v \in J$. The corollary therefore follows from the slope bounds of Propositions 6.4.1. □

6.5. **Small slope forms are classical.**

6.5.1. *Fargues' degree function.* We now recall some results on the degree of quasi-finite flat group schemes, following the papers [Far10, Far11]. Let K be a complete valued extension of $\mathbf{Q}_p$ with corresponding valuation $v : K \to \mathbf{R} \cup \{\infty\}$, which we assume to be normalized so that $v(p) = 1$. We also write $v : \mathcal{O}_K/p\mathcal{O}_K \to [0,1]$ for the induced map. If M is a finitely presented torsion $\mathcal{O}_K$-module, then we can write $M \cong \oplus_{i=1}^r \mathcal{O}_K/x_i$ for some $x_i \in \mathcal{O}_K$, and we set $\deg M := \sum_{i=1}^r v(x_i \mod p)$.

If H is a group scheme over $\mathcal{O}_K$, we let ω_H denote the conormal sheaf to the identity section. If H is finite flat, then ω_H is finitely presented and torsion over $\mathcal{O}_K$, and following Fargues we define the degree of H to be

$$\deg H := \deg \omega_H.$$

More generally, let $A \to A'$ be an isogeny of semi-abelian schemes with associated p-divisible groups $\mathcal{G} \to \mathcal{G}'$ over some analytic adic space S. We denote by $\omega_{\mathcal{G}}$ and $\omega_{\mathcal{G}'}$ the conormal sheaves of A and A' along their unit sections and by $\omega_{\mathcal{G}}^+$ and $\omega_{\mathcal{G}'}^+$ the subsheaf of integral differentials (which means that locally on S they arise from differentials on a formal model of A or A'). Let H be the kernel of $\mathcal{G} \to \mathcal{G}'$. This is a quasi-finite group scheme. To this isogeny we may attach a section δ_H of the locally free sheaf of rank one $\det\omega_{\mathcal{G}'} \otimes \det\omega_{\mathcal{G}}^{-1}$. Moreover, this section lies in the subsheaf $(\det\omega_{\mathcal{G}'})^+ \otimes (\det\omega_{\mathcal{G}}^{-1})^+$ of integral differential forms. For each point $x \in S$, we may compute the associated norm $|\delta_H|_x$, by choosing a trivialization of $(\det\omega_{\mathcal{G}'})^+ \otimes (\det\omega_{\mathcal{G}}^{-1})^+$ in a neighbourhood of x and viewing δ_H as a function (the norm $|\delta_H|_x$ is independent of the trivialization). If $x \in S$ is a rank one point with associated valuation normalized by $v_x(p) = 1$, and if $H_x \subset \mathcal{G}_x$ extends to a finite flat group scheme on a formal model $\mathfrak{G}_x$ of $\mathcal{G}_x$ over $\operatorname{Spec} k(x)^+$, then $v_x(\delta_H) = \deg H_x$.

6.5.2. *Neighbourhoods of the ordinary locus.* Recall from §4.3 that we define

$$K_p(I) = \prod_{v\in I} \mathrm{Kli}(v) \prod_{v\in I^c} \mathrm{Iw}(v).$$

We can consider $X_{K_p(I)K^p}$. Let $\mathcal{X}_{K_pK_p(I)}$ be the associated analytic space. For each $v|p$, we have an isogeny $\mathcal{G} \to \mathcal{G}'$ whose kernel is a quasi-finite group scheme H_v which is of order p away from the boundary. For each $v \in I^c$, we have an isogeny $\mathcal{G} \to \mathcal{G}'$ whose kernel is a quasi-finite group scheme L_v which is of order p^2 away from the boundary.

Let $\mathcal{X}^{rk1}_{K^pK_p(I)}$ be the subset of rank one points. To each rank one point x is associated a rank one valuation $v_x : \mathcal{O}_{\mathcal{X}_{K_p(I)K^p,x}} \to \mathbf{R} \cup \{\infty\}$ which we normalize

by $v_x(p) = 1$. If $v \in I$, we define $\deg_v : \mathcal{X}^{rk1}_{K^pK_p(I)} \to [0,1]$ by $\deg_v(x) = v_x(\delta_{H_v})$. Similarly, for all $v \in I^c$, we define $\deg_v : \mathcal{X}^{rk1}_{K^pK_p(I)} \to [0,2]$ by $\deg_v(x) = v_x(\delta_{L_v})$.

We can put all these degree functions together into a function

$$\deg : \mathcal{X}^{rk1}_{K^pK_p(I)} \to [0,1]^I \times [0,2]^{I^c}.$$

For each rational interval $J \subset [0,1]^I \times [0,2]^{I^c}$, there is a unique quasi-compact open subset $\mathcal{X}_{K^pK_p(I)}(J) \subset \mathcal{X}_{K^pK_p(I)}$ such that $\mathcal{X}_{K^pK_p(I)}(J)^{rk1} = \deg^{-1}(J)$.

Of particular interest is the multiplicative locus:

$$\mathcal{X}^{\mathrm{mult}}_{K^pK_p(I)} = \mathcal{X}_{K^pK_p(I)}(\{1\}^I \times \{2\}^{I^c}).$$

Let $(\epsilon_v) \in ([0,1]^I \times [0,2]^{I^c}) \cap \mathbf{Q}^{S_p}$ and set

$$\mathcal{X}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}) = \mathcal{X}_{K^pK_p(I)}(\prod_{v\in I}[1-\epsilon_v, 1]^I \times \prod_{v\in I^c}[2-\epsilon_v, 2]^{I^c}).$$

Observe that $\mathcal{X}^{\mathrm{mult}}_{K^pK_p(I)} = \mathcal{X}_{K^pK_p(I)}((0)_{v\in S_p})$ while $\{\mathcal{X}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p})\}_{\epsilon_v\to 0^+, \forall v\in S_p}$ is a fundamental system of strict neighbourhoods of $\mathcal{X}^{\mathrm{mult}}_{K^pK_p(I)}$.

All these spaces are stable under the action of $\mathcal{O}_{F_{(p)}}^{\times,+}$ on the polarization, and descend to open subspaces of $\mathcal{X}^{G_1}_{K^pK_p(I)}$. We can therefore add a superscript G_1 to any of these spaces with the obvious meaning.

6.5.3. *Comparison between spaces of overconvergent cohomology.* In this section we make the connection between the spaces $\mathcal{X}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p})$ (with $\epsilon_v \in ([0,1]^I \times [0,2]^{I^c}) \cap \mathbf{Q}^{S_p}$) that we just introduced and the spaces $\mathcal{X}_{K(I,p^n)}((\epsilon_v)_{v\in S_p})$ (say for parallel $n \in \mathbf{Z}_{\geq 1}$ and with $\epsilon_v \in ([0, n-\frac{1}{p-1}] \cap \mathbf{Q})^{S_p}$) introduced in §6.1.4. Both types of spaces are neighbourhoods of the multiplicative locus in an appropriate sense. The previous spaces are well adapted to the construction of interpolation sheaves and eigenvarieties while these new spaces appear naturally when one wants to prove classicity theorems.

There is a natural forgetful map $\mathcal{X}_{K(I,p^n)}((\epsilon_v)_{v\in S_p}) \to \mathcal{X}_{K^pK_p(I)}$. By [Pil20, Lem. 14.1.1] (for the places $v \in I$, and a trivial extension for the places $v \in I^c$), this map factors into a map $\mathcal{X}_{K(I,p^n)}((\epsilon_v)_{v\in S_p}) \to \mathcal{X}_{K_p(I)K^p}((1-\frac{2}{n}(n-\epsilon_v+\frac{1}{p-1}))_{v\in I} \times (2-\frac{2}{n}(n-\epsilon_v+\frac{1}{p-1}))_{v\in I^c})$. Observe that when $\epsilon_v = n-\frac{1}{p-1}$ and $n \to \infty$, $1-\frac{2}{n}(n-\epsilon_v+\frac{1}{p-1}) \to 1$ and $2-\frac{2}{n}(n-\epsilon_v+\frac{1}{p-1}) \to 2$. Conversely, by [Pil20, Lem. 14.1.2], there is a natural inclusion: $\mathcal{X}_{K_p(I)K^p}((\epsilon_v)_{v\in S_p}) \hookrightarrow \mathcal{X}_{K(I,p)}((1-\frac{1}{p-1})_{v\in S_p})$ for all $\epsilon_v \geq 1-\frac{1}{p}$ if $v \in I$ and $\epsilon_v \geq 2-\frac{1}{p}$ if $v \in I^c$.

Lemma 6.5.4. *Let $(\epsilon_v) \in ([1-\frac{1}{p}, 1)^I \times [2-\frac{1}{p}, 2)^{I^c}) \cap \mathbf{Q}^{S_p}$. Let κ be a classical algebraic weight.*

1. *The cohomology $\mathrm{R}\Gamma(\mathcal{X}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}), \omega^\kappa)$ carries an action of the operators $U_{v,1}$ and $U_{v,2}$.*
2. *The operator $\prod_{v\in I} U_{v,2} \prod_{v\in I^c} U_{v,1}$ is compact $\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}), \omega^\kappa)$.*
3. *The canonical map $\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}), \omega^\kappa) \to \mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K(I,p)}((1-\frac{1}{p-1})_{v\in S_p}), \omega^\kappa)$ induces a quasi-isomorphism on the finite slope part for $\prod_{v\in I} U_{v,2} \prod_{v\in I^c} U_{v,1}$.*
4. *The same holds for cuspidal cohomology.*

Proof. The definition of the operators is a routine computation. To prove compactness, we need to show that the operators improve convergence. This is entirely

parallel to Lemma 6.3.14. For the degree functions considered here this follows from of [Pil11, Prop. 2.3.6]. The quasi-isomorphism follows from an easy analytic continuation argument (see Lemma 6.5.18 below, for example). □

It is sometimes convenient to consider the dagger space

$$\mathcal{X}^{\mathrm{mult},\dagger}_{K^pK_p(I)} := \lim_{\epsilon_v \to 0^+} \mathcal{X}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}). \tag{6.5.5}$$

and its G_1-variant. In view of the previous lemma we can define the complex $e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))$ as being equal to $e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p(I)}((\epsilon_v)_{v\in S_p}), \omega^\kappa)$.

Lemma 6.5.6. *The complex $e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))$ is a perfect complex supported in degrees $[0, \#I]$.*

Proof. That the cohomology vanishes outside of degrees $[0, \#I]$ follows as usual by pushing forward to the minimal compactification. The finiteness of the cohomology follows from the compactness of U^I. □

6.5.7. *Main classicity theorem.* We now state our main classicity result for overconvergent cohomology, which we will prove using a generalization of the analytic continuation method of [Kas06] to higher degree cohomology, which was proved in [Pil20, §3]. Let $\kappa = (k_v, l_v)_{v|p}$ be a dominant algebraic weight. There is a canonical restriction map

$$\mathrm{R}\Gamma(\mathcal{X}_{K_p(I)K^p}, \omega^\kappa(-D)) \to \mathrm{R}\Gamma(\mathcal{X}^{\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))$$

which is equivariant for the Hecke operators $U_{v,1}$ and $U_{v,2}$.

Theorem 6.5.8. *The canonical map*

$$\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D))[U_{v,2} < k_v + l_v - 3 \ v \in I, \ U_{v,1} < l_v - 3 \ v \in I^c]$$

$$\to \mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))[U_{v,2} < k_v + l_v - 3 \ v \in I, \ U_{v,1} < l_v - 3 \ v \in I^c]$$

is a quasi-isomorphism.

Remark 6.5.9. The meaning of $[U_{v,2} < k_v + l_v - 3 \ v \in I, \ U_{v,1} < l_v - 3 \ v \in I^c]$ in Theorem 6.5.8 is the obvious one: it means the part of slope less than $k_v + l_v - 3$ for $U_{v,2}$ at $v \in I$ and less than $l_v - 3$ for $U_{v,1}$ at $v \in I^c$. (Note that while the individual operators $U_{v,1}, U_{v,2}$ do not act compactly on the complex on the right hand side, their product U does by Lemma 6.5.4 It follows the individual operators $U_{v,1}, U_{v,2}$ act compactly on the part of the complex with bounded slope for U, and so this small slope part is well-defined by the procedure explained at the start of this section.)

Remark 6.5.10. When $I = \emptyset$, Theorem 6.5.8 (for H^0) is proved in [BPS16]. It may be possible to improve on the bound $l_v - 3$ at the places $v \in I^c$, but this does not matter for our purposes.

6.5.11. *Hecke correspondences again.* Let $w \in I$. We consider the following correspondence, whose corresponding Hecke operator is $U^n_{w,2}$ (the nth iterate of $U_{w,2}$): $t_{w,n,1}, t_{w,n,2} : C^{(n)}_w \to \mathcal{X}_{K_p(I)K^p}$, which parametrizes $(\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)$ where the isogeny $\mathcal{G} \to \mathcal{G}_n$ has kernel $M_{n,w} \subset \mathcal{G}_w[p^{2n}]$ which is totally isotropic

and locally isomorphic to $(\mathcal{O}_{F_w}/p^n)^2 \oplus \mathcal{O}_{F_w}/p^{2n}$, and satisfies $M_{n,w} \cap H_w = \{0\}$. The first projection is

$$t_{w,n,1}\big((\mathcal{G}, \{H_v\}_{v\in I}, \{H_w \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)\big) = (\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c})$$

and the second projection is

$$t_{w,n,2}\big((\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)\big) = (\mathcal{G}_n, \{H'_v\}_{v\in I}, \{H'_v \subset L'_v\}_{v\in I^c})$$

where $\{H'_v\}_{v\in I}$ and $\{H'_v \subset L'_v\}_{v\in I^c}$ are the images of $\{H_v\}_{v\in I}$ and $\{H_v \subset L_v\}_{v\in I^c}$ in $\mathcal{G}_n$.

There are cohomological correspondences

$$(t_{w,n,1})_* t^*_{w,n,2}\omega^\kappa \to \omega^\kappa,\ (t_{w,n,1})_* t^*_{w,n,2}\omega^\kappa(-D) \to \omega^\kappa(-D),$$

which give $U^n_{w,2}$. Moreover, these cohomological correspondences restrict to

$$(t_{w,n,1})_* t^*_{w,n,2}(\omega^\kappa)^{++} \to p^{-3n}(\omega^\kappa)^{++},\ (t_{w,n,1})_* t^*_{w,n,2}(\omega^\kappa(-D))^{++} \to p^{-3n}(\omega^\kappa(-D))^{++},$$

and they induce maps on cohomology in the usual way.

Let $w \in I^c$. We consider the correspondence: $t_{w,n,1}, t_{w,n,2} : C^{(n)}_w \to \mathcal{X}_{K_p(I)K^p}$ which parametrizes $(\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)$ where the isogeny $\mathcal{G} \to \mathcal{G}_n$ has kernel $M_{n,w} \subset \mathcal{G}_w[p^n]$ which is totally isotropic, locally isomorphic to $(\mathcal{O}_{F_w}/p^n)^2$, and satisfies $M_{n,w} \cap L_w = \{0\}$. The first projection is

$$t_{w,n,1}\big((\mathcal{G}, \{H_v\}_{v\in I}, \{H_w \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)\big) = (\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c})$$

and the second projection is

$$t_{w,n,2}\big((\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_n)\big) = (\mathcal{G}_n, \{H'_v\}_{v\in I}, \{H'_v \subset L'_v\}_{v\in I^c})$$

where H'_v and $H'_v \subset L'_v$ are the images of H_v and $H_v \subset L_v$ in $\mathcal{G}_n$.

The Hecke operator attached to this correspondence is $U^n_{w,1}$ (the nth iterate of $U_{w,1}$). More precisely, there are cohomological correspondences

$$(t_{w,n,1})_* t^*_{w,n,2}\omega^\kappa \to \omega^\kappa,\ (t_{w,n,1})_* t^*_{w,n,2}\omega^\kappa(-D) \to \omega^\kappa(-D).$$

Moreover, these cohomological correspondences restrict to

$$(t_{w,n,1})_* t^*_{w,n,2}(\omega^\kappa)^{++} \to p^{-3n}(\omega^\kappa)^{++},\ (t_{w,n,1})_* t^*_{w,n,2}(\omega^\kappa(-D))^{++} \to p^{-3n}(\omega^\kappa(-D))^{++},$$

and they induce maps on cohomology.

Lemma 6.5.12. *Let $w \in I$. Let $x = (\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_1) \in C^{(1)}_w(\mathrm{Spa}(K, \mathcal{O}_K))$.*

(1) If $v \in I$ and $v \neq w$, we have $\deg H_v = \deg H'_v$.
(2) If $v \in I^c$, we have $\deg L_v = \deg L'_v$.
(3) We have $\deg H'_w \geq \deg H_w$, and in case of equality, $\deg H_w \in \{0, 1\}$.
(4) $\deg H'_w = 1 - \deg M_{1,w}/M_{1,w}[p]$.
(5) $\deg M_{1,w}[p]/pM_{1,w} = 1$, and $\deg pM_{1,w} \geq \deg M_{1,w}/M_{1,w}[p]$.
(6) Let $\epsilon \geq 0$. If $\deg M_w \leq 3 - 2\epsilon$, then $\deg H'_w \geq \epsilon$.

Proof. Parts (1) and (2) follow because the maps $H_v \to H'_v$, $L_v \to L'_v$ are isomorphisms. The remaining parts are [Pil20, Lem. 14.3.1, Cor. 14.3.1]. □

Lemma 6.5.13. *Let $w \in I^c$. Let $x = (\mathcal{G}, \{H_v\}_{v\in I}, \{H_v \subset L_v\}_{v\in I^c}, \mathcal{G} \to \mathcal{G}_1) \in C^{(1)}_w(\mathrm{Spa}(K, \mathcal{O}_K))$.*

(1) If $v \in I$, we have $\deg H_v = \deg H'_v$.
(2) If $v \in I^c$ and $v \neq w$, we have $\deg L_v = \deg L'_v$,

(3) We have $\deg L'_w \geq \deg L_w$*, and in case of equality,* $\deg L_w \in \{0,1,2\}$*.*
(4) $\deg L'_w = 2 - \deg M_{1,w}$*.*

Proof. Parts (1) and (2) follow as in Lemma 6.5.12. Parts (3) and (4) follow from [Pil11, Prop. 2.3.1, 2.3.2, Lem. 2.3.4] (and their proofs). □

Corollary 6.5.14. *Let* $w \in I^c$*. Let* $1 > \epsilon' \geq \epsilon > 0$*. There exists* $n \in \mathbf{Z}_{\geq 0}$ *such that for all intervals* $\prod_{v \neq w} J_v \subset [0,1]^I \times [0,2]^{I^c \setminus \{w\}}$*,*

$$U_{w,1}^n(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon, 2])) \subset \mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon', 2])$$

Proof. This follows from Lemma 6.5.13 (3) and the maximum principle; see [Pil11, Prop. 2.3.6]. □

Corollary 6.5.15. *Let* $w \in I$*. Let* $1 > \epsilon' \geq \epsilon > 0$*. There exists* $n \in \mathbf{Z}_{\geq 0}$ *such that for all intervals* $\prod_{v \neq w} J_v \subset [0,1]^{I \setminus \{w\}} \times [0,2]^{I^c}$*,*

$$U_{w,1}^n(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [\epsilon, 1])) \subset \mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [\epsilon', 1])$$

Proof. This follows in the same way as Corollary 6.5.14, using Lemma 6.5.12 (3). □

6.5.16. *First analytic continuation result.* Let $J = \prod_{v|p} J_v \subset [0,1]^I \times [0,2]^{I^c}$ be a product of intervals.

Lemma 6.5.17. *Let* $w \in I^c$*. Assume that* $J_w = [2-\epsilon, 2]$*. The operator* $U_{w,1}$ *acts on* $H^i(\mathcal{X}_{K^pK_p(I)}(J), \omega^\kappa)$*.*

Proof. In view of Lemma 6.5.13 (3), the correspondence $C_w^{(1)}$ restricts to

$$t_{w,1,2} : C_w^{(1)} \times_{t_{w,1,1}, \mathcal{X}_{K^pK_p(I)}} \mathcal{X}_{K^pK_p(I)}(J) \to \mathcal{X}_{K^pK_p(I)}(J). \qquad \square$$

We denote by $H^i(\mathcal{X}_{K_p(I)K^p}(J), \omega^\kappa)^{fs-U_{w,1}}$ the finite slope subspace for $U_{w,1}$. This is the subspace generated by classes which are annihilated by a polynomial in $U_{w,1}$ with non-zero constant term.

Lemma 6.5.18. *For all* $1 > \epsilon' \geq \epsilon > 0$*, the restriction map*

$$H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v \times [2-\epsilon', 2]), \omega^\kappa)^{fs-U_{w,1}} \to H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v \times [2-\epsilon, 2]), \omega^\kappa)^{fs-U_{w,1}}$$

is an isomorphism.

Proof. Take n as in Corollary 6.5.14. Let $f \in H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v \times [2-\epsilon, 2]), \omega^\kappa)^{fs-U_{w,1}}$ be a cohomology class. Let $P(X) = X^m + a_{m-1}X^{m-1} + \cdots + a_0$ be a polynomial with $a_0 \neq 0$ such that $P(U_{w,1})f = 0$. Therefore, if we set $Q(X) = -a_0^{-1}(P(X) - a_0)$, we obtain that $Q(U_{w,1})f = f$. By iteration we get that $Q(U_{w,1})^n f = f$. The operator

$$Q(U_{w,1})^n : H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v \times [2-\epsilon, 2]), \omega^\kappa) \to H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v \times [2-\epsilon, 2]), \omega^\kappa)$$

can be factored into:

$$H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon,2]),\omega^\kappa)\overset{\widetilde{Q(U_{w,1})^n}}{\to} H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon',2]),\omega^\kappa)$$
$$\overset{res}{\to} H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon,2]),\omega^\kappa),$$

where the map $\widetilde{Q(U_{w,1})^n}$ is the one coming from Corollary 6.5.14. We therefore get an extension $\tilde{f}$ of f to

$$H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon',2]),\omega^\kappa)$$

by setting $\tilde{f}=\widetilde{Q(U_{w,1})^n}f$. This proves the surjectivity of the map of the corollary.

We now prove injectivity. Let $f,g\in H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon',2]),\omega^\kappa)^{fs-U_{w,1}}$ be two classes having the same restriction to $H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon,2]),\omega^\kappa)^{fs-U_{w,1}}$. We can find a polynomial P as before such that $P(U_{w,1})f=P(U_{w,1})g=0$. Therefore, using the same notation as before, we get that $Q(U_{w,1})f=f$ and $Q(U_{w,1})g=g$. We can factor the operator $Q(U_{w,1})^n$ into:

$$H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon',2]),\omega^\kappa)\overset{res}{\to} H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon,2]),\omega^\kappa)$$
$$\overset{\widetilde{Q(U_{w,1})^n}}{\to} H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[2-\epsilon',2]),\omega^\kappa)$$

Since $res(f)=res(g)$, we deduce that $f=g$. □

The following two lemmas are the analogue of Lemma 6.5.18 for a place $w\in I$. The proofs are identical and left to the reader.

Lemma 6.5.19. *Let $w\in I$. Assume that $J_w=[1-\epsilon,1]$. The operator $U_{w,2}$ acts on $H^i(\mathcal{X}_{K^pK_p(I)}(J),\omega^\kappa)$.*

We denote by $H^i(\mathcal{X}_{K_p(I)K^p}(J),\omega^\kappa)^{fs-U_{w,2}}$ the finite slope subspace. This is the subspace generated by classes which are annihilated by a polynomial in $U_{w,2}$ with non-zero constant term.

Lemma 6.5.20. *For all $1>\epsilon\geq\epsilon'>0$, the restriction map*

$$H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[1-\epsilon,1]),\omega^\kappa)^{fs-U_{w,2}}\to H^i(\mathcal{X}_{K^pK_p(I)}(\prod_{v\neq w} J_v\times[1-\epsilon',1]),\omega^\kappa)^{fs-U_{w,2}}$$

is an isomorphism.

6.5.21. *More analytic continuation results.* Let $w\in I^c$. Let $0<\epsilon\leq 1$. The cohomological correspondences:

$$(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])})\to\omega^\kappa|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])}$$

and

$$(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,2])})\to\omega^\kappa|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,2])}$$

can be related if we work with torsion coefficients.

Proposition 6.5.22. *Let $0 < \epsilon < \epsilon'$. There is a factorization of the Hecke correspondence $U^n_{w,1}$:*

$$\begin{tikzcd}
 & (\omega^\kappa/p^{n(l_w(1-\epsilon')-3)}(\omega^\kappa)^{++})|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])} \\
(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])}) \arrow[ur] \arrow[dr] & \\
 & (\omega^\kappa/p^{n(l_w(1-\epsilon')-3)}(\omega^\kappa)^{++})|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,2])} \arrow[uu] \\
(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,2])}) \arrow[uu] \arrow[ur] &
\end{tikzcd}$$

Proof. Let $x \in \mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]))$. We have to find a neighbourhood U of x and to construct a canonical map

$$\begin{aligned}
&(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])})(t^{-1}_{w,n,1}(U)) \\
&\to (\omega^\kappa/p^{n(l_w(1-\epsilon')-3)}(\omega^\kappa)^{++})|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,2])}(U).
\end{aligned}$$

Pick $\epsilon'' \in (\epsilon, \epsilon')$ such that for all $y \in t^{-1}_{w,n,1}(x)$ corresponding to a subgroup $M_{w,n} \subset \mathcal{G}$, we have $|\delta_{M_{w,n}}|_y \neq |p^{n(1-\epsilon'')}|_y$. It follows that there exists an open neighbourhood U of x and a disjoint decomposition $t^{-1}_{w,n,1}(U) = V \coprod W$, such that for all $y \in W$, we have $|\delta_{M_{w,n}}|_y > |p^{n(1-\epsilon'')}|_y$, and for all $y \in V$, we have $|\delta_{M_{w,n}}|_y < |p^{n(1-\epsilon'')}|_y$.

The cohomological correspondence is a map:

$$t^*_{w,n,2}(\omega^\kappa)^{++}(V) \oplus t^*_{w,n,2}(\omega^\kappa)^{++}(W) \to \omega^\kappa(U).$$

The image of $t^*_{w,n,2}(\omega^\kappa)^{++}(V)$ lands in $p^{n(l_v(1-\epsilon'')-3)}(\omega^\kappa)^{++}(U)$ (see [Pil20, Lem. 14.6.1]). Therefore, we have a factorization:

$$t^*_{w,n,2}(\omega^\kappa)^{++}(t^{-1}_{w,n,1}(U)) \to t^*_{w,n,2}(\omega^\kappa)^{++}(W) \to \omega^\kappa/p^{n(l_v(1-\epsilon'')-3)}(\omega^\kappa)^{++}(U).$$

On the other hand, we claim that $t_{w,n,2}(W) \subset \mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon, 2]))$.

Indeed, let $x' \in U$ and let $y' \in t^{-1}_{w,n,1}(\{x\})$. Without loss of generality, we may assume that x' and y' are rank one points. Let us define $M_{w,i} = M_{w,n}|_{y'}[p^i]$ for all $1 \leq i \leq n$. Then we have a sequence of isogenies:

$$\mathcal{G}_w|_{x'} \to \mathcal{G}_w|_{x'}/M_{w,1} \to \cdots \to \mathcal{G}_w|_{x'}/M_{w,n}$$

We let $L_{w,i}$ be the image of $L_w|_{x'}$ in $\mathcal{G}_w/M_{w,i}$ (so that $L_{w,n} = L_w|_{t_{w,n,2}(y')}$). Then, by Lemma 6.5.13 (4), we have that $\deg L_{w,n} = 2 - \deg M_{w,n}/M_{w,n-1}$. On the other hand, for all $1 \leq i \leq n-1$, the map $p : M_{w,i+1}/M_{w,i} \to M_{w,i}/M_{w,i-1}$ is a generic isomorphism. It follows that $\deg M_{w,i+1}/M_{w,i} \leq \deg M_{w,i}/M_{w,i-1}$. Since $\deg M_{w,n} = \sum_{i=1}^n \deg M_{w,i}/M_{w,i-1}$ and $\deg M_{w,n} \leq n(1-\epsilon'')$, we deduce that $\deg M_{w,n}/M_{w,n-1} \leq 1 - \epsilon''$ and therefore $\deg L_{w,n} \geq 1 + \epsilon'' > 1 + \epsilon$, as required.

We can therefore produce the expected map as the composition:

$$(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])})(t_{w,n,1}^{-1}(U))$$

$$\to (t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2])})(W) = t_{w,n,2}^*(\omega^\kappa)^{++}(W)$$

$$\to \omega^\kappa/p^{n(l_v(1-\epsilon')-3)}(\omega^\kappa)^{++}(U). \qquad \square$$

Corollary 6.5.23. *Let $P = X^m + a_{m-1}X^{m-1} + \cdots + a_0$ be a polynomial, with the property that all the roots a of P satisfy $v(a) < l_w - 3$. Then there is a map*

$$\mathrm{ext} : H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),\omega^\kappa)[P(U_{w,1}) = 0] \to$$

$$\left(\lim_n H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa/p^n(\omega^\kappa)^{++})\right)[P(U_{w,1}) = 0]$$

such that the composite of ext *followed by restriction to $\mathcal{X}_{K_p(I)Kp}(\prod_{v\neq w} J_v \times [1+\epsilon,2])$ is the natural map induced by $\omega^\kappa \to \omega^\kappa/p^n(\omega^\kappa)^{++}$.*

Furthermore, the composite of the restriction map

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa)[P(U_{w,1}) = 0] \to$$

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),\omega^\kappa)[P(U_{w,1}) = 0]$$

followed by ext *is the natural map induced by $\omega^\kappa \to \omega^\kappa/p^n(\omega^\kappa)^{++}$.*

Proof. Let $\epsilon' > 0$ be such that for all roots a of P, we have $l_w(1-\epsilon')-3 > v(a)$. Let $\alpha = \inf_a\{l_w(1-\epsilon')-3-v(a)\}$ (so that in particular $\alpha > 0$). By Lemma 6.5.18, we can assume that $0 < \epsilon < \epsilon'$. Suppose that $f \in H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),\omega^\kappa)$ satisfies $P(U_{w,1})f = 0$. By rescaling f, we can and do also assume that $f \in H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2]),(\omega^\kappa)^{++})$. Let $Q(X) = -a_0^{-1}(P(X)-a_0)$ so that $Q(U_{w,1})f = f$.

Since $Q(U_{w,1})$ can be written as a sum of products of the $\frac{1}{a}U_{w,1}$, where a runs over the roots of P, it follows from Proposition 6.5.22 that the map

$$Q(U_{w,1})^n : H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2]),(\omega^\kappa)^{++}) \to H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1+\epsilon,2]),\omega^\kappa)$$

$$\to H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),(\omega^\kappa)/p^{n\alpha}(\omega^\kappa)^{++})$$

can actually be factored into:

$$\begin{tikzcd}
H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),(\omega^\kappa)^{++}) \arrow[dr, "Q(\widetilde{U_{w,1}})^n"] \arrow[dd] & \\
& H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),(\omega^\kappa)/p^{n\alpha}(\omega^\kappa)^{++}) \arrow[dl] \\
H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),(\omega^\kappa)/p^{n\alpha}(\omega^\kappa)^{++}) &
\end{tikzcd}$$

We define sections $f_n \in H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa/p^{n\alpha}(\omega^\kappa)^{++})$ by $f_n = Q(\widetilde{U_{w,1}})^n(f)$. It follows from the definitions that $f_n = f_{n-1}$ in

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa/p^{(n-1)\alpha-m(l_w(1-\epsilon')-\alpha)}(\omega^\kappa)^{++})$$

and that $Q(U_{w,1})f_n = f_n$ in

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa/p^{(n\alpha-m(l_w(1-\epsilon')-\alpha)}(\omega^\kappa)^{++})$$

(see the proof of [Pil20, Cor. 14.6.1] for a similar verification). We let $\mathrm{ext}(f)$ be the projective system given by the f_n.

It remains to check that if f is the restriction of a class in

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),\omega^\kappa),$$

then the f_n are obtained from the natural map $\omega^\kappa \to \omega^\kappa/p^n(\omega^\kappa)^{++}$. This follows easily from the factorization

$$\begin{tikzcd}
H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),(\omega^\kappa)^{++}) \arrow[dr, "Q(U_{w,1})^n"] \arrow[dd] & \\
& H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,2]),(\omega^\kappa)/p^{n\alpha}(\omega^\kappa)^{++}) \\
H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [1+\epsilon,2]),(\omega^\kappa)/p^{n\alpha}(\omega^\kappa)^{++}) \arrow[ur, "Q(\widetilde{U_{w,1}})^n"'] &
\end{tikzcd}$$

□

The next proposition and corollary are the analogue of the above results for a place $w \in I$. The proofs are virtually identical to the above, and are left to the reader (or look at [Pil20, Prop. 14.6.1, Cor. 14.6.1]).

Proposition 6.5.24. *Let $w \in I$, and let $0 < \epsilon < \epsilon'$. There is a factorization of the Hecke correspondence $U^n_{w,2}$:*

$$\begin{tikzcd}
 & (\omega^\kappa/p^{n(l_w+k_w-3-2\epsilon' k_v)}(\omega^\kappa)^{++})|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[\epsilon,1])} \\
(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[\epsilon,1])}) \arrow[ur] \arrow[dr] & \\
 & (\omega^\kappa/p^{n(l_w+k_w-3-2\epsilon' k_w)}(\omega^\kappa)^{++})|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[0,1])} \arrow[uu] \\
(t_{w,n,1})_*(t_{w,n,2})^*((\omega^\kappa)^{++}|_{\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v\times[1,0])}) \arrow[uu] \arrow[ur] &
\end{tikzcd}$$

Corollary 6.5.25. *Let $w \in I$, and let $1 > \epsilon \geq 0$. Let $P = X^m + a_{m-1}X^{m-1} + \cdots + a_0$ be a polynomial, with the property that all the roots a of P satisfy $v(a) < k_w + l_w - 3$. Then there is a map*

$$\mathrm{ext} : H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [\epsilon,1]), \omega^\kappa)[P(U_{w,2}) = 0] \to$$

$$\left(\lim_n H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,1]), \omega^\kappa/p^n(\omega^\kappa)^{++})\right)[P(U_{w,2}) = 0]$$

such that the composite of ext *followed by restriction to $\mathcal{X}_{K_p(I)Kp}(\prod_{v\neq w} J_v \times [\epsilon,1])$ is the natural map induced by $\omega^\kappa \to \omega^\kappa/p^n(\omega^\kappa)^{++}$.*

Furthermore, the composite of the restriction map

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [0,1]), \omega^\kappa)[P(U_{w,2}) = 0] \to$$

$$H^i(\mathcal{X}_{K_p(I)K^p}(\prod_{v\neq w} J_v \times [\epsilon,1]), \omega^\kappa)[P(U_{w,2}) = 0]$$

followed by ext *is the natural map induced by $\omega^\kappa \to \omega^\kappa/p^n(\omega^\kappa)^{++}$.*

6.5.26. *Proof of the main classicality theorem.* Let $S \subset S_p$ be a subset. Let $J(S,\epsilon) = \prod_{v\in S\cap I}[0,1] \times \prod_{v\in S\cap I^c}[0,2] \times \prod_{S^c\cap I}[\epsilon,1] \times \prod_{S^c\cap I^c}[1+\epsilon,2]$. We say that a cohomology class $f \in H^i(\mathcal{X}_{K_p(I)K^p}(J(S,\epsilon),\omega^\kappa)$ is of finite slope if for all $v|p$, there is a polynomial P_v all of whose roots are nonzero, such that:

- if $v \in I^c$, $P_v(U_{v,1})f = 0$,
- if $v \in I$, $P_v(U_{v,2})f = 0$.

Lemma 6.5.27. *The canonical map*

$$H^i(\mathcal{X}^{G_1}_{K_p(I)K^p}(J(S,\epsilon)),\omega^\kappa) \to \lim H^i(\mathcal{X}^{G_1}_{K_p(I)K^p}(J(S,\epsilon)),\omega^\kappa/p^n(\omega^\kappa)^{++})$$

is surjective and induces an isomorphism on the finite slope part.

Proof. The surjectivity follows from [Pil20, Prop. 3.2.1]. The injectivity can be proved in exactly the same way as [Pil20, Lem. 14.7.1]. We have put the superscript G_1 because we need some finiteness property to deduce the injectivity. □

Lemma 6.5.28. *Choose polynomials P_v such that*

- *if $v \in I^c$, all the roots a of P_v satisfy $v(a) < l_v - 3$, and*
- *if $v \in I$, all the roots a of P_v satisfy $v(a) < k_v + l_v - 3$.*

Write $U_v = U_{v,1}$ if $v \in I^c$, and $U_v = U_{v,2}$ if $v \in I$. If $S \subset T$, then the natural restriction map

$$\begin{array}{c} H^i(\mathcal{X}^{G_1}_{K_p(I)K^p}(J(T,\epsilon)),\omega^\kappa)[P_v(U_v)=0]_{v\in S_p} \\ \Big\downarrow \\ H^i(\mathcal{X}^{G_1}_{K_p(I)K^p}(J(S,\epsilon)),\omega^\kappa)[P_v(U_v)=0]_{v\in S_p} \end{array}$$

is an isomorphism.

Proof. By induction, it is enough to treat the case $T = S \cup \{w\}$ for some w. The result then follows from Lemma 6.5.27 (applied to both S and T), together with Corollary 6.5.23 and Corollary 6.5.25. □

Proof of Theorem 6.5.8. This follows immediately from Lemma 6.5.28, applied with the choices $S = \emptyset$ and $T = S_p$. □

6.6. **Application to ordinary cohomology.** In this section we study the case $\#I = 1$, where we are able to relate the Hida complexes constructed in §4 to the overconvergent cohomology considered in this section. Our first result is the following, which shows in particular that in this case the ordinary classes in H^1 are overconvergent. The proof can be viewed as a generalization of the familiar argument for GL_2 which shows that ordinary p-adic modular forms are overconvergent (see [BT99, Lem. 1]), by using the continuity of the ordinary projector to the finite-dimensional space of ordinary forms.

Recall that we defined the complex M_I in Theorem 4.6.1. By Theorem 4.6.1 (3), for all classical algebraic weights κ with $l_v = 2$ for $v \in I$ and $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$ we have

$$M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p = e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult}}_{K_p(I)K^p},\omega^\kappa(-D)).$$

Proposition 6.6.1. *Suppose that $\#I = 1$. For all classical algebraic weights κ with $l_v = 2$ for $v \in I$ and $k_v \equiv l_v \equiv 2 \pmod{p-1}$ for all $v|p$, the restriction map*

$$e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p},\omega^\kappa(-D)) \to M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p$$

induces an injective map on H^0 and a surjective map on H^1.

Proof. The injectivity of the map on H^0s is clear. In the case $F = \mathbf{Q}$, the surjectivity of the map on H^1s is proved in [Pil20, Lem. 14.8.2]; we now recall this argument in our setting. Let $\pi : \mathcal{X}_{K_p(I)K^p} \to \mathcal{X}^*_{K_p(I)K^p}$ be the projection to the minimal compactification; as usual, we have $\mathrm{R}^i\pi_*\omega^\kappa(-D) = 0$ for $i > 0$. Let $\mathcal{X}^{*,\mathrm{mult}}_{K_p(I)K^p}$ be the image of $\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}$ (the rigid analytic generic fibre of $\mathfrak{X}^I_{K_p(I)K^p}$) in the minimal compactification; it admits an affinoid cover $\mathcal{X}^{*,\mathrm{mult}}_{K_p(I)K^p} = U_1 \cup U_2$.

Then the complex $\mathrm{R}\Gamma(\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$ is represented by the complex

$$H^0(U_1, \omega^\kappa(-D)) \oplus H^0(U_2, \omega^\kappa(-D)) \to H^0(U_1 \cap U_2, \omega^\kappa(-D)).$$

The terms of this complex are Banach spaces; a norm giving their topology is provided by taking an appropriate formal model. The topology on $H^1(\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$ is the induced quotient topology, which coincides with the topology obtained by declaring that $H^1(\mathfrak{X}^I_{K_p(I)K^p}, \omega^\kappa(-D))$ is open and bounded.

The complex $\mathrm{R}\Gamma(\mathcal{X}^{\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))$ is represented by the subcomplex of overconvergent sections

$$H^0(U_1, \omega^{\kappa,\dagger}(-D)) \oplus H^0(U_2, \omega^{\kappa,\dagger}(-D)) \to H^0(U_1 \cap U_2, \omega^{\kappa,\dagger}(-D)),$$

where by definition for an open U,

$$\omega^{\kappa,\dagger}(U) := \varinjlim_{V \supset U} \omega^\kappa(V)$$

where V runs over the strict neighbourhoods of U.

It follows in particular that the map

$$H^1(\mathcal{X}^{\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D)) \to H^1(\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$$

has dense image. The operator U^I is continuous on $H^1(\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$ (consider the action of U^I on $H^1(\mathfrak{X}^I_{K_p(I)K^p}, \omega^\kappa(-D))$), so the projection

$$H^1(\mathcal{X}^{G_1,\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D)) \to e(U^I)H^1(\mathcal{X}^{G_1,\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$$

is also continuous (we have introduced the superscript G_1 to make sure that the projector $e(U^I)$ is well defined, the passage from the cohomology of $\mathcal{X}^{\mathrm{mult}}_{K_p(I)K^p}$ to the cohomology of $\mathcal{X}^{G_1,\mathrm{mult}}_{K_p(I)K^p}$ is given by a projector so all density statements are preserved). It follows that the induced map

$$e(U^I)H^1(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D)) \to e(U^I)H^1(\mathcal{X}^{G_1,\mathrm{mult}}_{K_p(I)K^p}, \omega^\kappa(-D))$$

has dense image. But the target is a finite-dimensional Banach space over $\mathbf{C}_p$, so its topology is the unique one extending that on $\mathbf{C}_p$, and in particular it contains no proper dense subspaces, so we are done. □

Proposition 6.6.2. *Suppose that $\#I \leq 1$. For all classical algebraic weights κ with $l_v = 2$ for $v \in I$, we have the equality of Euler characteristics:*

$$EC(e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))) = EC(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p).$$

Proof. By Theorem 4.6.1 and Lemma 6.5.6, both complexes are perfect complexes in degrees $[0, 1]$. By Corollary 6.4.3 and Proposition 6.3.7, we have that

$$EC(e(U^I)H^i_{cusp}(G_1, \dagger, \kappa)) \leq EC(e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D)))$$

and the inequality is an equality if $k_v - l_v \geq 3$ for all $v|p$. By Proposition 6.6.1, we have that

$$EC(e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))) \leq EC(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p).$$

Consequently, it suffices to prove that

$$EC(e(U^I)H^i_{cusp}(G_1, \dagger, \kappa)) \geq EC(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p).$$

By Theorem 6.3.16, and Theorem 4.6.1, both Euler characteristics under consideration are locally constant functions of κ. It therefore suffices to prove the statement when $l_v \geq C$ for all $v \in I^c$, and $k_v - l_v \geq C$ for all $v|p$. In this range of weights we can compare these cohomology to classical cohomology.

It follows from Theorem 6.5.8 that

$$e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D))) = e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D))).$$

We claim that the natural map

$$\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p}, \omega^\kappa(-D))) \to \mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D)))$$

induces a quasi-isomorphism

$$e(T^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p}, \omega^\kappa(-D))) \to e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D))).$$

Indeed, by Theorem 3.10.1 (2), the cohomology groups on each side can be computed in terms of automorphic representations, and the claim follows from Proposition 2.4.26 as explained in Remark 6.6.3 below.

Now, it follows from Theorem 4.6.1 that the map $e(T^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p}, \omega^\kappa(-D)) \to M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p$ is an isomorphism on H^0 and is injective on H^1. Putting this all together, the proposition follows. □

Remark 6.6.3. Let us point out a subtle point in the proof of Proposition 6.6.2. In order to use Proposition 2.4.26 one needs to check that for any $v|p$, any representation π_v of $\mathrm{GSp}_4(\mathcal{O}_{F_v})$ contributing to either $e(T^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K^pK_p}, \omega^\kappa(-D)))$ or $e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D)))$ is ordinary. For all places $v \in I^c$, this is true essentially by definition since the two Hecke operators at $v|p$ occur in the projector. For places $v \in I$, this is a bit more subtle since only one operator T_v or $U_{v,2}$ is involved in the definition of the projector. The $U_{v,2}$-ordinarity of a local representation π_v with Hecke parameters

$$[\alpha_v p^{1-k_v/2}, \beta_v p^{1-k_v/2}, \beta_v^{-1} p^{k_v/2}, \alpha_v^{-1} p^{k_v/2}]$$

implies that $\alpha_v\beta_v$ is a p-adic unit. Ordinarity means that α_v and β_v are both p-adic units. This is implied by $U_{v,2}$-ordinarity if we assume the Katz–Mazur inequality which says the the Newton polygon is above the Hodge polygon with the same initial and terminal point. Indeed, in our case, the Katz–Mazur inequality translates into the condition that α_v and β_v are p-adic integers.

However, this inequality is subtle at non-cohomological weights. For $F = \mathbf{Q}$ the Katz–Mazur inequality for H^0 and H^1 classes is proved in [Pil20, Prop. 14.9.1], and the argument generalizes without difficulty to our case. We also remark that for classes in the H^0, we can use eigenvarieties to deduce that the Katz–Mazur inequality holds in non-cohomological weights because it holds at cohomological weights. A similar argument will apply for classes in the H^1 once eigenvarieties are constructed for H^1 cohomology classes. Note that alternatively we could force the

Katz–Mazur inequality by localizing further at certain p-adically integral eigenvalues of the operators $T_{v,1}$ and $U_{\mathrm{Kli}(v),1}$, and in fact such a localization will be in force in the rest of the paper. We could also directly deduce the Katz–Mazur inequality for classes in the H^1 from the corresponding inequality for classes in the H^0 after making a non-Eisenstein localization (because after making such a localization, the Euler characteristic vanishes). Such a localization will also be in force in the rest of the paper. In view of this, we do not spell out the details of the generalization of [Pil20, Prop. 14.9.1] to our setting.

Theorem 6.6.4. *Suppose that $\#I \leq 1$. For all classical algebraic weights κ with $l_v = 2$ for $v \in I$, we have a canonical isomorphism:*

$$e(U^I)\mathrm{R}\Gamma(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D)) = M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p$$

Proof. For all classical algebraic weights, let us denote by $d_i(\kappa)$ the dimension of $H^i(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p)$ and by $d_i^\dagger(\kappa)$ the dimension of $e(U^I)H^i_{cusp}(G_1, \dagger, \kappa)$. We deduce from Proposition 6.6.2 that $d_1(\kappa) - d_0(\kappa) = d_1^\dagger(\kappa) - d_0^\dagger(\kappa)$ for all κ. Therefore $d_1(\kappa) - d_1^\dagger(\kappa) = d_0(\kappa) - d_0^\dagger(\kappa)$ for all κ. By Proposition 6.6.1, the first difference is non-positive and the second difference is non-negative, so both are equal to zero. We deduce in particular that the map $e(U^I)H^i_{cusp}(G_1, \dagger, \kappa) \to H^i(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p)$ is an isomorphism.

We now consider the composite

$$e(U^I)H^i(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K_p(I)K^p}, \omega^\kappa(-D)) \to e(U^I)H^i_{cusp}(G_1, \dagger, \kappa) \to H^i(M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{C}_p).$$

By Proposition 6.6.1 this composite map is an isomorphism for $i = 0$, and is surjective for $i = 1$. On the other hand, the first map is injective for $i = 1$ by Corollary 6.4.3, and we have just seen that the second map is an isomorphism. It follows that the composite is injective for $i = 1$, and is thus an isomorphism, as required. □

Finally, we deduce the following classicity theorem.

Theorem 6.6.5. *Suppose that $\#I \leq 1$, and that κ is a classical algebraic weight with $l_v = 2$ if $v \in I$, and $l_v \geq 4$ if $v \notin I$. Then the canonical map*

$$e(U^I)\mathrm{R}\Gamma(X^{G_1}_{K_p(I)K^p}, \omega^\kappa(-D))[1/p] \to M_I \otimes^{\mathbf{L}}_{\Lambda_I,\kappa} \mathbf{Q}_p$$

is a quasi-isomorphism.

Proof. This follows from Theorem 6.6.4 and Theorem 6.5.8. □

7. The Taylor–Wiles/Calegari–Geraghty method

In this section, we implement the Taylor–Wiles patching method to patch the complexes M_I constructed in §4. More precisely, we carry out the analogue of the patching argument using "balanced modules" which was introduced in [CG18], and used there to study weight one modular forms for $\mathrm{GL}_2/\mathbf{Q}$. This argument works in situations where the cohomology appears in at most two degrees, which for us means that $\#I \leq 1$; we are restricted to working in this case due to the limitations of our understanding of the cohomology of our complexes in higher degree, as was the case in §4 and §6. For our modularity result, it is crucial to be able to work with $I = S_p$; we will do this in §8 by considering the spaces of modular forms coming from the various complexes with $\#I \leq 1$.

The papers [GT05, Pil12, CG20] apply the Taylor–Wiles method to GSp_4 over $\mathbf{Q}$, but a number of changes are needed in order to apply it over general totally real fields. We do not attempt to prove results in maximal generality, but instead develop the minimal amount of material that we need. The reader familiar with the literature on modularity lifting theorems will not find many surprises, but we highlight a few things that may be less standard:

- In §7.3, we study the ordinary deformation rings at places dividing p. We show that their generic fibres are irreducible under a rather mild p-distinguishedness assumption; in particular, this assumption is not sufficient to guarantee that the deformation rings are formally smooth, and it takes us some effort to prove the irreducibility. Working in this generality is important for our applications to modularity of abelian surfaces in §10. For the potential modularity results of §9, however, it would be enough to work with a stronger p-distinguishedness assumption which would guarantee the formal smoothness.
- In §7.4, we prove the statements about local deformation rings needed for Taylor's "Ihara avoidance argument"; the proofs are similar to those for GL_n, although there are some complications which arise because the relationship between conjugacy classes and characteristic polynomials is more complicated. We also need to do some additional work to handle the case $p = 3$; again, this is crucial for §10, although it is not needed for §9.
- In §7.5, we study the "big image" conditions needed in the Taylor–Wiles method. Here our approach is slightly different from that of [CHT08] and the papers that followed it; again, this is with the applications of §10 in mind, where it is important to be able to consider representations with image $\mathrm{GSp}_4(\mathbf{F}_3)$. For the same reason, when we impose a condition at an auxiliary prime which will allow us to assume that our Shimura varieties are at neat level, we make the weakest hypothesis that we can, at the expense of slightly complicating the corresponding local representation theory.
- We make repeated use of the doubling results of §5; they are needed in order to prove local-global compatibility for the Galois representations we consider, and also to compare the spaces of p-adic modular forms for different I.
- Our implementation of the "Ihara avoidance" argument of [Tay08] uses the framework of [EG14, Sho18], and compares the underlying cycles of various patched modules. In particular, we use the patched modules with $I = \emptyset$ to prove a local result, which we then apply to the patched modules with $\#I = 1$. In order to apply Ihara avoidance, we repeatedly use the fact that the Galois representation associated to our abelian surface is pure, to deduce that the corresponding points on the generic fibres of the local deformation rings are smooth; we use this smoothness to be able to compute the dimensions of various spaces of p-adic automorphic forms, using a characteristic 0 version of the freeness arguments of Diamond and Fujiwara [Dia97]. While we do not use the full strength of purity, since we make arguments with base change we would otherwise need to impose a hypothesis of being "stably generic" on our local Galois representations, and we do not know of any natural examples where this condition is known, but purity is not.

Having carried out the patching argument, we know from the results of §6 that for each I with $\#I \leq 1$ there is a nonzero space of ordinary p-adic modular forms corresponding to our given Galois representation, which are "overconvergent in the direction of I". We will combine these spaces in §8, using as an input that by a version of Diamond's multiplicity one argument [Dia97], we know the dimensions of these spaces when $\#I \leq 1$ (they are given by the expected product of local terms). (Here we are again using our assumption that the local Galois representations are pure, in order to know that the corresponding points of the generic fibres of the local deformation rings are smooth. A similar characteristic zero version of Diamond's argument first appeared in [All16].)

7.1. **Galois deformation rings.** We let E be a finite extension of $\mathbf{Q}_p$ with ring of integers $\mathcal{O}$, uniformizer λ and residue field k. We will always assume that E is chosen to be large enough such that all irreducible components of all deformation rings that we consider, and all irreducible components of their special fibres, are geometrically irreducible. (We are always free to enlarge E in all of the arguments that we make, so this is not a serious assumption.) Given a complete Noetherian local $\mathcal{O}$-algebra Λ with residue field k, we let CNL_Λ denote the category of complete Noetherian local Λ-algebras with residue field k. We refer to an object in CNL_Λ as a CNL_Λ-algebra.

We fix a totally real field F, and let S_p be the set of places of F above p. We assume that E contains all embeddings of F into an algebraic closure of E. We also fix a continuous absolutely irreducible homomorphism $\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$. We assume throughout that $p > 2$.

Let S be a finite set of finite places of F containing S_p and all places at which $\overline{\rho}$ is ramified. We write F_S for the maximal subextension of $\overline{F}/F$ which is unramified outside S, and write $G_{F,S}$ for $\mathrm{Gal}(F_S/F)$. For each $v \in S$, we fix $\Lambda_v \in \mathrm{CNL}_\mathcal{O}$, and set $\Lambda = \widehat{\otimes}_{v \in S} \Lambda_v$, where the completed tensor product is taken over $\mathcal{O}$. Then CNL_Λ is a subcategory of CNL_{Λ_v} for each $v \in S$, via the canonical map $\Lambda_v \to \Lambda$.

Remark 7.1.1. In our applications, we will take $\Lambda_v = \mathcal{O}$ if $v \nmid p$. If $v|p$, then we will take Λ_v to be an Iwasawa algebra.

Fix a character $\psi : G_{F,S} \to \Lambda^\times$ lifting $\nu \circ \overline{\rho}$.

Definition 7.1.2. A *lift*, also called a *lifting*, of $\overline{\rho}|_{G_{F_v}}$ is a continuous homomorphism $\rho : G_{F_v} \to \mathrm{GSp}_4(A)$ to a CNL_{Λ_v}-algebra A, such that $\rho \bmod \mathfrak{m}_A = \overline{\rho}|_{G_{F_v}}$ and $\nu \circ \rho = \psi|_{G_{F_v}}$.

We let $\mathcal{D}_v^\square$ denote the set-valued functor on CNL_{Λ_v} that sends A to the set of lifts of $\overline{\rho}|_{G_{F_v}}$ to A. This functor is representable (see for example [Bal12, Thm. 1.2.2]), and we denote the representing object by $R_v^\square$.

Let $x \in \operatorname{Spec} R_v^\square[1/p]$ be a closed point. By [Tay08, Lem. 1.6] the residue field of x is a finite extension E'/E. Let $\rho_x : G_{F_v} \to \mathrm{GL}_n(E')$ be the corresponding specialization of the universal lifting. By an argument of Kisin, $(R_v^\square[1/p])_x^\wedge$ is the universal lifting ring for ρ_x, i.e. if A is an Artinian local E'-algebra with residue field E' and if $\rho : \Gamma \to \mathrm{GSp}_4(A)$ is a continuous representation lifting ρ_x, then there is a unique continuous map of E'-algebras $(R_v^\square[1/p])_x^\wedge \to A$ so that the universal lift pushes forward to ρ. (See [All16, Thm. 1.2.1] for the analogous result for GL_n; the result for GSp_4 can be proved by an identical argument.) We say that x is *smooth* if $(R_v^\square[1/p])_x^\wedge$ is regular. Let $\mathrm{ad}^0 \rho_x$ denote the Lie algebra $\mathfrak{g}^0(E')$ with the

adjoint action of G_F via ρ_x; then we have the following convenient criterion for x to be smooth.

Lemma 7.1.3. *Suppose that $v \nmid p$. Then the point x is smooth if and only if $(\mathrm{ad}^0 \rho_x)(1)^{G_{F_v}} = 0$. In particular, if ρ_x is pure, then x is smooth.*

Proof. The first claim is a special case of [BG19, Cor. 3.3.4, Rem. 3.3.6]. If ρ_x is pure, then $\mathrm{Hom}_{E'[G_{F_v}]}(\rho_x, \rho_x(1)) = 0$ (because the definition of purity is easily seen to preclude the existence of a morphism between the corresponding Weil–Deligne representations), as required. □

Definition 7.1.4. A *local deformation problem* for $\overline{\rho}|_{G_{F_v}}$ is a subfunctor $\mathcal{D}_v$ of $\mathcal{D}_v^{\square}$ satisfying the following:

- $\mathcal{D}_v$ is represented by a quotient R_v of $R_v^{\square}$.
- For all $A \in \mathrm{CNL}_{\Lambda_v}$, $\rho \in \mathcal{D}_v(A)$, and $a \in \ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}_4(k))$, we have $a\rho a^{-1} \in \mathcal{D}_v(A)$.

Definition 7.1.5. A *global deformation problem* is a tuple

$$\mathcal{S} = (\overline{\rho}, S, \{\Lambda_v\}_{v \in S}, \psi, \{\mathcal{D}_v\}_{v \in S}),$$

where:

- $\overline{\rho}$, S, $\{\Lambda_v\}_{v \in S}$ and ψ are as above.
- For each $v \in S$, $\mathcal{D}_v$ is a local deformation problem for $\overline{\rho}|_{G_{F_v}}$.

As in the local case, a *lift* (or *lifting*) of $\overline{\rho}$ is a continuous homomorphism $\rho : G_{F,S} \to \mathrm{GSp}_4(A)$ to a CNL_Λ-algebra A, such that $\rho \bmod \mathfrak{m}_A = \overline{\rho}$ and $\rho \circ \nu = \psi$. We say that two lifts $\rho_1, \rho_2 : G_{F,S} \to \mathrm{GSp}_4(A)$ are *strictly equivalent* if there is an $a \in \ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}_4(k))$ such that $\rho_2 = a\rho_1 a^{-1}$. A *deformation of* $\overline{\rho}$ is a strict equivalence class of lifts of $\overline{\rho}$.

For a global deformation problem

$$\mathcal{S} = (\overline{\rho}, S, \{\Lambda_v\}_{v \in S}, \psi, \{\mathcal{D}_v\}_{v \in S}),$$

we say that a lift $\rho : G_F \to \mathrm{GSp}_4(A)$ is of *type* $\mathcal{S}$ if $\rho|_{G_{F_v}} \in \mathcal{D}_v(A)$ for each $v \in S$. Note that if ρ_1 and ρ_2 are strictly equivalent lifts of $\overline{\rho}$, and ρ_1 is of type $\mathcal{S}$, then so is ρ_2. A *deformation of type* $\mathcal{S}$ is a strict equivalence class of lifts of type $\mathcal{S}$, and we denote by $\mathcal{D}_{\mathcal{S}}$ the set-valued functor that takes a CNL_Λ-algebra A to the set of lifts $\rho : G_F \to \mathrm{GSp}_4(A)$ of type $\mathcal{S}$.

Given a subset $T \subseteq S$, a *T-framed lift of type* $\mathcal{S}$ is a tuple $(\rho, \{\gamma_v\}_{v \in T})$, where ρ is a lift of type $\mathcal{S}$, and $\gamma_v \in \ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}_4(k))$ for each $v \in T$. We say that two T-framed lifts $(\rho_1, \{\gamma_v\}_{v \in T})$ and $(\rho_2, \{\gamma'_v\}_{v \in T})$ to a CNL_Λ-algebra A are strictly equivalent if there is an $a \in \ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}_4(k))$ such that $\rho_2 = a\rho_1 a^{-1}$, and $\gamma'_v = a\gamma_v$ for each $v \in T$. A strict equivalence class of T-framed lifts of type $\mathcal{S}$ is called a *T-framed deformation of type* $\mathcal{S}$. We denote by $\mathcal{D}_{\mathcal{S}}^T$ the set valued functor that sends a CNL_Λ-algebra A to the set of T-framed deformations to A of type $\mathcal{S}$.

The functors $\mathcal{D}_{\mathcal{S}}$ and $\mathcal{D}_{\mathcal{S}}^T$ are representable (as we are assuming that $\overline{\rho}$ is absolutely irreducible), and we denote their representing objects by $R_{\mathcal{S}}$ and $R_{\mathcal{S}}^T$ respectively. If T is empty, then $R_{\mathcal{S}} = R_{\mathcal{S}}^T$, and otherwise the natural map $R_{\mathcal{S}} \to R_{\mathcal{S}}^T$ is formally smooth of relative dimension $11\#T - 1$. Indeed $\mathcal{D}_{\mathcal{S}}^T \to \mathcal{D}_{\mathcal{S}}$ is a torsor under $(\prod_{v \in T} \widehat{\mathrm{GSp}_4})/\widehat{\mathbf{G}_m}$. Define $\mathcal{T}$ to be the coordinate ring of $(\prod_{v \in T} \widehat{\mathrm{GSp}_4})/\widehat{\mathbf{G}_m}$ over Λ. This is a power series algebra over Λ in $11\#T - 1$ variables.

Lemma 7.1.6. *The choice of a representative $\rho_{\mathcal{S}}\colon G_F \to \mathrm{GSp}_4(R_{\mathcal{S}})$ for the universal type $\mathcal{S}$ deformation determines a splitting of the torsor $\mathcal{D}_{\mathcal{S}}^T \to \mathcal{D}_{\mathcal{S}}$ and a canonical isomorphism $R_{\mathcal{S}}^T \cong R_{\mathcal{S}} \widehat{\otimes}_\Lambda \mathcal{T}$.*

Proof. This is obvious. □

7.2. **Galois cohomology and presentations.** Fix a global deformation problem

$$\mathcal{S} = (\overline{\rho}, S, \{\Lambda_v\}_{v\in S}, \psi, \{\mathcal{D}_v\}_{v\in S}),$$

and for each $v \in S$, let R_v denote the object representing $\mathcal{D}_v$. Let T be a subset of S containing S_p, with the property that $\Lambda_v = \mathcal{O}$ and $\mathcal{D}_v = \mathcal{D}_v^\square$ for all $v \in S \smallsetminus T$. Define $R_{\mathcal{S}}^{T,\mathrm{loc}} = \widehat{\otimes}_{v\in T} R_v$, with the completed tensor product being taken over $\mathcal{O}$. It is canonically a Λ-algebra, via the canonical isomorphism $\widehat{\otimes}_{v\in T}\Lambda_v \cong \widehat{\otimes}_{v\in S}\Lambda_v$. For each $v \in T$, the morphism $\mathcal{D}_{\mathcal{S}}^T \to \mathcal{D}_v$ given by $(\rho, \{\gamma_v\}_{v\in T}) \mapsto \gamma_v^{-1}\rho|_{G_{F_v}}\gamma_v$ induces a local Λ_v-algebra morphism $R_v \to R_{\mathcal{S}}^T$. We thus have a local Λ-algebra morphism $R_{\mathcal{S}}^{T,\mathrm{loc}} \to R_{\mathcal{S}}^T$.

The relative tangent space of this map is computed by a standard calculation in Galois cohomology, which we now recall. We let $\mathrm{ad}\,\overline{\rho}$ (resp. $\mathrm{ad}^0\,\overline{\rho}$) denote $\mathfrak{g}(k)$ (resp. $\mathfrak{g}^0(k)$), with the adjoint G_F-action via $\overline{\rho}$.

The trace pairing $(X, Y) \mapsto \mathrm{tr}(XY)$ on $\mathrm{ad}^0\,\overline{\rho}$ is perfect and G_F-equivariant, so $\mathrm{ad}^0\,\overline{\rho}(1)$ is isomorphic to the Tate dual of $\mathrm{ad}^0\,\overline{\rho}$. We define

$$H^1_{\mathcal{S},T}(\mathrm{ad}^0\,\overline{\rho}) := \ker\left(H^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}) \to \prod_{v\in T} H^1(F_v, \mathrm{ad}^0\,\overline{\rho})\right),$$

$$H^1_{\mathcal{S}^\perp,T}(\mathrm{ad}^0\,\overline{\rho}(1)) := \ker\left(H^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) \to \prod_{v\in S\smallsetminus T} H^1(F_v, \mathrm{ad}^0\,\overline{\rho}(1))\right).$$

Proposition 7.2.1. *Continue to assume that T contains S_p, and that for all $v \in S \smallsetminus T$ we have $\Lambda_v = \mathcal{O}$ and $\mathcal{D}_v = \mathcal{D}_v^\square$. Then there is a local Λ-algebra surjection $R_{\mathcal{S}}^{T,\mathrm{loc}}[[X_1, \dots, X_g]] \to R_{\mathcal{S}}^T$, with*

$$g = h^1_{\mathcal{S}^\perp,T}(\mathrm{ad}^0\,\overline{\rho}(1)) - h^0(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) - \sum_{v|\infty} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}) + \sum_{v\in S\setminus T} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}(1)) + \#T - 1.$$

Proof. We follow [Kis09, §3.2]. By [Kis09, Lem. 3.2.2] (or rather the same statement for GSp_4, which has an identical proof), the claim of the proposition holds with

$$g = h^1_{\mathcal{S},T}(\mathrm{ad}^0\,\overline{\rho}) - h^0(F_S/F, \mathrm{ad}\,\overline{\rho}) + \sum_{v\in T} h^0(F_v, \mathrm{ad}\,\overline{\rho}).$$

By [DDT97, Thm. 2.19] (and the assumption that $\overline{\rho}$ is absolutely irreducible, which implies that $h^0(F_S/F, \mathrm{ad}\,\overline{\rho}^0) = 0$), we have

$$h^1_{\mathcal{S},T}(\mathrm{ad}^0\,\overline{\rho}) = h^1_{\mathcal{S}^\perp,T}(\mathrm{ad}^0\,\overline{\rho}(1)) - h^0(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) - \sum_{v|\infty} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}) + \sum_{v\in S\setminus T} h^1(F_v, \mathrm{ad}^0\,\overline{\rho}) - \sum_{v\in S} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}).$$

The result follows from the local Euler characteristic formula and Tate local duality. □

7.3. **Local deformation problems, $l = p$.** Assume from now on that p splits completely in F. Let v be a place of F lying over p. If $x \in k^\times$, then we write $\lambda_x : G_{F_v} \to k^\times$ for the unramified character with $\lambda_x(\mathrm{Frob}_v) = x$.

Definition 7.3.1. We say that $\overline{\rho}|_{G_{F_v}}$ is *p-distinguished weight* 2 *ordinary* if it is conjugate to a representation of the form

$$\begin{pmatrix} \lambda_{\overline{\alpha}_v} & 0 & * & * \\ 0 & \lambda_{\overline{\beta}_v} & * & * \\ 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}_v}^{-1} & 0 \\ 0 & 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}_v}^{-1} \end{pmatrix},$$

where $\overline{\alpha}_v \neq \overline{\beta}_v$.

If $\overline{\rho}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary, then we say that a lift $\rho : G_{F_v} \to \mathrm{GSp}_4(\mathcal{O})$ of $\overline{\rho}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary if ρ itself is conjugate to a representation of the form

$$\begin{pmatrix} \lambda_{\alpha_v} & 0 & * & * \\ 0 & \lambda_{\beta_v} & * & * \\ 0 & 0 & \varepsilon^{-1}\lambda_{\beta_v}^{-1} & 0 \\ 0 & 0 & 0 & \varepsilon^{-1}\lambda_{\alpha_v}^{-1} \end{pmatrix}$$

where α_v, β_v lift $\overline{\alpha}_v$, $\overline{\beta}_v$ respectively. Note that ρ is then automatically semistable, although not necessarily crystalline.

Remark 7.3.2. The terminology "weight 2 ordinary" is not ideal, but we were unable to find a better alternative. Possibilities include "P-ordinary" (referring to the Siegel parabolic subgroup), which clashes with "p-distinguished", or "semistable ordinary". We could of course restrict to the crystalline case and use "flat ordinary" representations, but as it costs us little to allow semistable representations, and it may prove to be useful in future applications, we have not done this.

Remark 7.3.3. For the purposes of proving the potential modularity of abelian surfaces, it would suffice to work with a stronger p-distinguishedness hypothesis, as in [CG20]. In particular, by assuming that none of $\overline{\alpha}_v^2, \overline{\beta}_v^2, \overline{\alpha}_v\overline{\beta}_v$ are equal to 1, we could arrange that the various deformation rings considered in this section are formally smooth. However, such a hypothesis is very restrictive in the case $p = 3$, and in particular would seriously restrict the applicability of our modularity lifting theorems to proving the modularity (as opposed to potential modularity) of particular abelian surfaces.

We assume from now on that $\overline{\rho}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary for all $v|p$; the roles of $\overline{\alpha}_v, \overline{\beta}_v$ in the definition of p-distinguished weight 2 ordinary are symmetric, and we fix a labelling of $\overline{\alpha}_v, \overline{\beta}_v$ for each $v|p$.

Set $\Lambda_{v,1} = \mathcal{O}[[\mathcal{O}_{F_v}^\times(p)]]$, $\Lambda_{v,2} = \mathcal{O}[[(\mathcal{O}_{F_v}^\times(p))^2]]$, where $\mathcal{O}_{F_v}^\times(p) = 1 + p\mathcal{O}_{F_v}$ denotes the pro-p completion of $\mathcal{O}_{F_v}^\times$. Both $\Lambda_{v,1}$ and $\Lambda_{v,2}$ are formally smooth over $\mathcal{O}$ (because we are assuming that $F_v = \mathbf{Q}_p$). Let Λ_v be either $\Lambda_{v,1}$ or $\Lambda_{v,2}$. There is a canonical character $I_{F_v} \to \mathcal{O}_{F_v}^\times(p)$ given by $\mathrm{Art}_{F_v}^{-1}$, and we define a pair of characters $\theta_{v,i} : I_{F_v} \to \Lambda_v$, $i = 1, 2$ as follows: if $\Lambda_v = \Lambda_{v,1}$, then we let $\theta_{v,1} =$

$\theta_{v,2} = \theta_v$ be the natural character and if $\Lambda_v = \Lambda_{v,2}$, then we let $\theta_{v,i}$ correspond to the embedding $\mathcal{O}_{F_v}^\times(p)$ to $(\mathcal{O}_{F_v}^\times(p))^2$ given by the ith copy.

Let $\bar{\mathfrak{r}}_v$ denote a choice of either $\overline{\alpha}_v$ or $\overline{\beta}_v$, and write $\bar{\mathfrak{r}}_v' := \overline{\alpha}_v\overline{\beta}_v/\bar{\mathfrak{r}}_v$ for the other choice. Recall that we have the Borel subgroup B of GSp_4 consisting of matrices of the form

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

We let P denote the subgroup of B consisting of matrices of the form

$$\begin{pmatrix} * & 0 & * & * \\ 0 & * & * & * \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}.$$

If $A \in \mathrm{CNL}_{\Lambda_v}$, then we say that a lift $\rho_A : G_{F_v} \to \mathrm{GSp}_4(A)$ of $\overline{\rho}|_{G_{F_v}}$ is $(B, \bar{\mathfrak{r}}_v)$-ordinary if there is an increasing filtration of free A-submodules

$$0 = \mathrm{Fil}^0 \subset \mathrm{Fil}^1 \subset \cdots \subset \mathrm{Fil}^4 = A^4$$

of A^4 by $A[G_{F_v}]$-submodules such that the action of G_{F_v} on $\mathrm{Fil}^i / \mathrm{Fil}^{i-1}$ is via a character χ_i with $\overline{\chi}_1 = \lambda_{\bar{\mathfrak{r}}_v}$, $\overline{\chi}_2 = \lambda_{\bar{\mathfrak{r}}_v'}$, $\chi_1|_{I_{F_v}} = \theta_{v,1}$, $\chi_2|_{I_{F_v}} = \theta_{v,2}$, and $\chi_3 = \varepsilon^{-1}\chi_2^{-1}$, $\chi_4 = \varepsilon^{-1}\chi_1^{-1}$.

By [CHT08, Lem. 2.4.6] such a filtration is unique; since $\{(\mathrm{Fil}^{4-i})^\perp\}$ gives another filtration satisfying the same conditions, we see that ρ_A is $\ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}(k))$-conjugate to a representation of the form

$$\begin{pmatrix} \chi_1 & * & * & * \\ 0 & \chi_2 & * & * \\ 0 & 0 & \varepsilon^{-1}\chi_2^{-1} & * \\ 0 & 0 & 0 & \varepsilon^{-1}\chi_1^{-1} \end{pmatrix}$$

where χ_1, χ_2 are as above.

If $\Lambda_v = \Lambda_{v,1}$ (so that $\theta_{v,1} = \theta_{2,v}$), then we say that ρ_A is P-ordinary if it is both $(B, \overline{\alpha}_v)$-ordinary and $(B, \overline{\beta}_v)$-ordinary; equivalently, if ρ_A is $\ker(\mathrm{GSp}_4(A) \to \mathrm{GSp}(k))$-conjugate to a representation of the form

$$\begin{pmatrix} \chi_1 & 0 & * & * \\ 0 & \chi_2 & * & * \\ 0 & 0 & \varepsilon^{-1}\chi_2^{-1} & 0 \\ 0 & 0 & 0 & \varepsilon^{-1}\chi_1^{-1} \end{pmatrix}.$$

If $\Lambda_v = \Lambda_{v,2}$ (resp. $\Lambda_v = \Lambda_{v,1}$) then we let $\mathcal{D}_v^{B,\bar{\mathfrak{r}}_v}$ (resp. $\mathcal{D}_v^P$) be the subfunctor of $(B, \bar{\mathfrak{r}}_v)$-ordinary lifts (resp. of P-ordinary lifts). By [CHT08, Lem. 2.4.6], we see that $\mathcal{D}_v^{B,\bar{\mathfrak{r}}_v}$ and $\mathcal{D}_v^P$ are local deformation problems in the sense of Definition 7.1.4, so they are represented by CNL_{Λ_v}-algebras $R_v^{B,\bar{\mathfrak{r}}_v}$, R_v^P respectively.

Most of the rest of this section is devoted to the proof of the following result.

Proposition 7.3.4. *The generic fibres $R_v^{B,\bar{\mathfrak{r}}_v}[1/p]$, $R_v^P[1/p]$ are irreducible, and are of relative dimensions* 16 *and* 14 *respectively over $\mathbf{Q}_p$.*

Our arguments are rather ad hoc, and will require a number of preliminary lemmas.

7.3.5. *Ordinary deformation rings for* GL_2. We begin by studying some ordinary deformation rings for GL_2. As well as being a warmup for our main arguments, we will often be able to show that our deformation rings for GSp_4 are formally smooth over a completed tensor product of deformation rings for GL_2, thus reducing to this case.

Let $\overline{r} : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(k)$ be of the form

$$\begin{pmatrix} \lambda_{\overline{\alpha}} & * \\ 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1} \end{pmatrix}.$$

Set $\Lambda = \mathcal{O}[[1 + p\mathbf{Z}_p]]$, and write $\theta : I_{\mathbf{Q}_p} \to \Lambda$ for the canonical character defined above. If $A \in \mathrm{CNL}_\Lambda$, then we say that a lift of $\overline{r}$ to $r : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(A)$ is ordinary if it is $\ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(k))$-conjugate to a representation of the form

$$\begin{pmatrix} \chi & * \\ 0 & \varepsilon^{-1}\chi^{-1} \end{pmatrix}$$

where $\overline{\chi} = \lambda_{\overline{\alpha}}$ and $\chi|_{I_{\mathbf{Q}_p}} = \theta$. As above, this is a local deformation problem, and is represented by a CNL_Λ-algebra $R^{B_2,\square}$, where B_2 denotes the Borel subgroup of GL_2 of upper triangular matrices.

We write

$$\overline{r} = \begin{pmatrix} \lambda_{\overline{\alpha}} & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1}\eta_{\alpha^2} \\ 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1} \end{pmatrix},$$

where η_{α^2} is a cocycle in $Z^1(\mathbf{Q}_p, \overline{\varepsilon}\lambda_{\overline{\alpha^2}})$. Rescaling our basis vectors has the effect of changing η_{α^2} by a coboundary, so we can and do think of η_{α^2} as a class in $H^1(\mathbf{Q}_p, \overline{\varepsilon}\lambda_{\overline{\alpha^2}})$.

Let $\mathfrak{b}_2$ be the Lie algebra of B_2, given by the matrices

$$\mathfrak{b}_2 = \begin{pmatrix} \nu + x_\alpha & x_{\alpha^2} \\ 0 & \nu - x_\alpha \end{pmatrix},$$

where $\mathrm{ad}^0_{B_2}$ corresponds to $\nu = 0$. With respect to the basis given by the matrices corresponding to the variables $\{x_{\alpha^2}, x_\alpha\}$ — that is, the basis $\left\{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\}$ — the Galois representation $\mathrm{ad}^0_{B_2}\,\overline{r}$ is given explicitly as follows:

$$\begin{pmatrix} \overline{\varepsilon}\lambda_{\overline{\alpha}^2} & -2\eta_{\alpha^2} \\ 0 & 1 \end{pmatrix}.$$

Note that if M is annihilated by p, then $H^2(\mathbf{Q}_p, M)$ is given by $H^0(\mathbf{Q}_p, M^*)^\vee \simeq \mathrm{Hom}_{G_{\mathbf{Q}_p}}(M, \overline{\varepsilon})^\vee$. It follows that $h^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}) = 0$ unless $\overline{\alpha}^2 = 1$ and $\eta_{\alpha^2} = 0$, in which case we have $h^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}) = 1$. We write R^{B_2} for $R^{B_2,\square}$ unless we particularly want to emphasize the GL_2-framing variables.

Lemma 7.3.6. *The generic fibre $R^{B_2}[1/p]$ is irreducible, and has relative dimension 5 over $\mathbf{Q}_p$.*

Proof. By a standard argument (see [Maz89, Prop. 2]), R^{B_2} has a presentation of the form $\mathcal{O}[[x_1, \ldots, x_r]]/(y_1, \ldots, y_s)$, where

$$r = 4 - h^0(\mathbf{Q}_p, \mathrm{ad}_{B_2}\,\overline{r}) + h^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}) = 3 - h^0(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}) + h^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}),$$
$$s = h^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}\,\overline{r}).$$

Note that, *a priori*, even when $s > 0$, some of the y_i may vanish, although one does not expect this to happen. By the local Euler characteristic formula, $r - s =$

$3 + \dim \mathrm{ad}^0_{B_2} \overline{r} = 5$. In particular, if $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2} \overline{r}) = 0$, then R^{B_2} is formally smooth over $\mathcal{O}$ of relative dimension 5, and we are done.

If $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2} \overline{r}) \neq 0$, then the above discussion shows that $\overline{\alpha} = \pm 1$, $\overline{r}$ is split, and $s = h^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2} \overline{r}) = 1$. Since any quotient of a formal power series ring by a single relation is a local complete intersection, it follows from the presentation of the previous paragraph that R^{B_2} is a local complete intersection, and in particular S_2. Note that, at this point, we don't know if the relation y_1 is non-zero or not, so we do not as yet know the dimension of R^{B_2}.

Twisting by a quadratic character, we can and do suppose that $\overline{\alpha} = 1$, so that $\overline{r} = 1 \oplus \overline{\varepsilon}^{-1}$. We begin by showing that $\operatorname{Spec} R^{B_2}[1/p]$ is connected, following [Ger19, Lem. 3.13]. Note that the map $\operatorname{Spec} R^{B_2}[1/p] \to \operatorname{Spec} \Lambda[1/p]$ admits a section, because we can always find a lift of the form $\chi \oplus \chi^{-1}\varepsilon^{-1}$. Since $\operatorname{Spec} \Lambda[1/p]$ is connected, it therefore suffices to show that the fibres of this map over closed points x of $\operatorname{Spec} \Lambda[1/p]$ are connected.

By (for example) the proof of [BLGGT14b, Lem. 1.2.2] (see also [BG19, Lem. 3.4.1]), the irreducible components of $\operatorname{Spec} R^{B_2}[1/p]$ are fixed by conjugation by elements of $\mathrm{GL}_2(R^{B_2})$ whose image in $\mathrm{GL}_2(k)$ is diagonal. It is therefore obviously the case that all the closed points which are conjugate to representations of the form $\chi \oplus \chi^{-1}\varepsilon^{-1}$ lie in the same connected component of the fibre over x, so it suffices to show that each closed point of the form

$$r = \begin{pmatrix} \chi & * \\ 0 & \varepsilon^{-1}\chi^{-1} \end{pmatrix}$$

lies in the same connected component as the corresponding point with $* = 0$. To this end, we consider the representation

$$r_t : \mathrm{diag}(t, t^{-1}) r \, \mathrm{diag}(t, t^{-1})^{-1} \to \mathrm{GL}_2(\mathcal{O}\langle t \rangle).$$

Note that the specializations of this representation at $t = 0$ and $t = 1$ correspond to the two closed points that we are considering.

Letting $A \subset \mathcal{O}\langle t \rangle$ be the closed subalgebra generated by the matrix entries of the elements of the image r_t, one checks exactly as in the proof of [BLGGT14b, Lem. 1.2.2] that A is a complete local Noetherian $\mathcal{O}$-algebra with residue field k. Since $\overline{r}$ is split, it follows that the representation r_t arises from a map $R^{B_2} \to A$. Since A is a domain (being a subring of $\mathcal{O}\langle t \rangle$), we see that the points corresponding to $t = 0$ and $t = 1$ lie on the same irreducible component, as required.

To see that $R^{B_2}[1/p]$ is moreover irreducible, it is enough to check that it is normal, or equivalently that it is R_1 and S_2. We have already seen that it is S_2, and to show that it is R_1, it suffices to show that there is an open regular subscheme U of $R^{B_2}[1/p]$ whose complement has codimension at least 2. We will in fact show that there is such a subscheme with the property that the tangent space at any closed point $u \in U$ has dimension 5, thus also proving the statement about the dimension of $R^{B_2}[1/p]$ (if the one relation in our presentation of R^{B_2} was trivial, then R^{B_2} would be formally smooth of relative dimension 6, and there would be no such points).

Over R^{B_2}, we have a universal lifting $r^{\mathrm{univ}} : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(R^{B_2})$, and we let $H^2_{\mathrm{ord}} := H^2(G_{\mathbf{Q}_p}, \mathrm{ad}^0_{B_2} r^{\mathrm{univ}})$, a finite R^{B_2}-module. Since the cohomology of $G_{\mathbf{Q}_p}$ vanishes in degree greater than 2, the formation of H^2 is compatible with specialization, so that in particular if x is a closed point of $R^{B_2}[1/p]$ with corresponding

representation $r_x : G_{\mathbf{Q}_p} \to \mathrm{GL}_2(E_x)$ (with E_x a finite extension of $\mathbf{Q}_p$), then $H^2(G_{\mathbf{Q}_p}, \mathrm{ad}^0_{B_2} r_x) = H^2_{\mathrm{ord}} \otimes_{R^{B_2}} E_x$.

We let U be the complement of the support of H^2_{ord} in $\operatorname{Spec} R^{B_2}[1/p]$. (It is not obvious *a priori* that U is not empty, but we will prove this below.) Then at any closed point $x \in U$, we have $H^2(G_{\mathbf{Q}_p}, \mathrm{ad}^0_{B_2} r_x) = 0$, so by a standard Galois cohomology calculation (essentially identical to the one used in the first paragraph of this proof), U is formally smooth over $\mathbf{Q}_p$ at x, with relative tangent space of dimension 5. It follows that U is regular.

The complement of U is the support of H^2, so just as above, its closed points are those x for which ρ_x is a direct sum of two characters whose ratio is the cyclotomic character. But in any Zariski open neighbourhood of such a point there are points of U (for example, points which are a direct sum of two characters whose ratio is not the cyclotomic character, given by twisting the characters occurring in ρ_x by unramified characters), so U is dense in $\operatorname{Spec} R^{B_2}[1/p]$, and $\operatorname{Spec} R^{B_2}[1/p]$ is equidimensional of relative dimension 5 over $\mathbf{Q}_p$.

It remains to show that the complement of U (that is, the support of H^2) has codimension at least 2, or equivalently that it has relative dimension at most 3 over $\mathbf{Q}_p$. In fact, it has relative dimension at most 2 over $\mathbf{Q}_p$: the only freedom we have is to make twists of the two characters in ρ_x (and the determinant is fixed), so the corresponding dimension is the dimension of GL_2 minus the dimension of the centralizer of $\begin{pmatrix} 1 & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}$, which equals $4 - 2 = 2$, so we are done. $\square$

Let $R^{B_2,\triangledown}$ denote the B_2-valued framed deformation ring of $\overline{r}$ with fixed determinant. It follows from the assumption that $\overline{r}$ is p-distinguished that $R^{B_2,\square}$ is formally smooth over $R^{B_2,\triangledown}$ of relative dimension

$$\dim \mathrm{ad}^0_{\mathrm{GL}_2} - \dim \mathrm{ad}^0_{B_2} = 3 - 2 = 1$$

(see Lemma 7.3.12 for the details of an analogous argument in the symplectic case).

It will prove useful to give (somewhat) explicit descriptions of $R^{B_2,\triangledown}$ (and thus R^{B_2}) in a number of explicit cases. Lemma 7.3.7 below will also give another proof of Lemma 7.3.6, although not one we shall generalize to the symplectic context.

Let $\gamma \in k$, and let

$$\overline{r} = \begin{pmatrix} \lambda_\gamma & \overline{\varepsilon}^{-1}\lambda_\gamma^{-1}\eta \\ 0 & \overline{\varepsilon}^{-1}\lambda_\gamma^{-1} \end{pmatrix}.$$

Via restriction to the character in the upper left hand corner, the ring $R^{B_2,\triangledown}$ is naturally an algebra over the universal deformation ring R^{GL_1} for GL_1. This gives a map

$$R^{\mathrm{GL}_1} \to R^{B_2,\triangledown}.$$

The ring R^{GL_1} is formally smooth of relative dimension 2 over $\mathcal{O}$, and also formally smooth over the Iwasawa algebra Λ corresponding to restricting the character to inertia. Let us choose isomorphisms $\Lambda = \mathcal{O}[[y_2]]$ and $R^{\mathrm{GL}_1} = \mathcal{O}[[y_1, y_2]]$. In the lemma below, we shall use y_i for the variables of $R^{B_2,\triangledown}$ corresponding to the algebra structure over R^{GL_1}, we use x_i for framing variables, and z_i for variables related to extensions (informally corresponding to the upper right corner). More precisely, by "framing variables" we mean the following: the map $c \mapsto (1 + \varepsilon c)\overline{r}$ gives an isomorphism

$$Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(R^{B_2,\triangledown}, k[\varepsilon]/(\varepsilon^2)) = \mathrm{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k),$$

where $\mathfrak{m}$ is the maximal ideal of $R^{B_2,\urcorner}$ and on the level of reduced tangent spaces, the framing variables are by definition the coboundaries $B^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \subset Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$.

Lemma 7.3.7. *Let $\mathfrak{m}$ denote the maximal ideal of $R^{B_2,\urcorner}$. The ring $R^{B_2,\urcorner}$ is a complete intersection, is flat over Λ, and is irreducible of relative dimension 4 over $\mathcal{O}$. The rings $R^{B_2,\urcorner}$ as R^{GL_1}-algebras have the following explicit presentations.*

(1) If $\gamma^2 \neq 1$ and $\eta \neq 0$, then $R^{B_2,\urcorner} \simeq \mathcal{O}[[x_1, x_2, y_1, y_2]]$.
(2) If $\gamma^2 \neq 1$ and $\eta = 0$, then $R^{B_2,\urcorner} \simeq \mathcal{O}[[x_1, z_1, y_1, y_2]]$.
(3) If $\gamma^2 = 1$ but $\eta \neq 0$, then:
  (a) $R^{B_2,\urcorner}$ is formally smooth over $\mathcal{O}$,
  (b) $R^{B_2,\urcorner}$ is formally smooth over Λ unless η is peu ramifiée,
  (c) $R^{B_2,\urcorner} \simeq \mathcal{O}[[x_1, x_2, z_1, y_1, y_2]]/g_\eta$, where $g_\eta = c_\eta y_1 + d_\eta y_2 \mod (\lambda, \mathfrak{m}^2)$ for $[c_\eta : d_\eta] \in \mathbf{P}^1(k)$, and where $[c_\eta : d_\eta]$ depends only on $\eta \in H^1(\overline{\varepsilon})$.
(4) If $\gamma^2 = 1$ and $\eta = 0$, then:
  (a) $R^{B_2,\urcorner} \simeq \mathcal{O}[[x_1, z_1, z_2, y_1, y_2]]/g$, where
$$g = z_1 y_1 + z_2 y_2 \mod (\lambda, \mathfrak{m}^3).$$
  (b) The special fibre $R^{B_2,\urcorner}/\lambda$ is not formally smooth.

Remark 7.3.8. Explicit descriptions of ordinary deformation rings (even over general extensions $K/\mathbf{Q}_p$) for GL_2 have been given by Böckle in [BÖ0, §7]. However, we require some precise information about these rings as algebras over R^{GL_1} and Λ which is not *explicitly* given in the required form in [BÖ0], and thus we have found it easier to give the argument below. However, all the methods below already appear (in a more complicated setting) in previous work of Böckle and others.

Proof. We first note that $\mathrm{ad}^0_{B_2}$ is simply the 2-dimensional representation given by
$$0 \to k(\lambda_\gamma^2 \overline{\varepsilon}) \to \mathrm{ad}^0_{B_2} \to k \to 0,$$
and where the extension class is given by η (so this is just a twist of $\overline{r}$). The framed tangent space has dimension
$$\dim Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = \dim H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) + \dim B^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}),$$
with precisely
$$\dim \mathrm{ad}^0_{B_2} - \dim H^0(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$$
framing variables. The maps from R^{GL_1} and from Λ correspond on tangent spaces to the maps $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, k)$ and $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(I_{\mathbf{Q}_p}, k)$ given by the composites of the maps
$$Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, k) \to H^1(I_{\mathbf{Q}_p}, k).$$
We now consider the four possible cases in turn.

If $\gamma^2 \neq 1$ and $\eta \neq 0$, then $H^0(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ is trivial, and there are two framing variables x_1 and x_2. The map from $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ to $H^1(\mathbf{Q}_p, k)$ is an isomorphism. Note that $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = 0$, and so $R^{B_2,\urcorner}$ is formally smooth over $\mathcal{O}$ and all statements are clear in this case.

If $\gamma^2 \neq 1$ and $\eta = 0$, then $H^0(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = k$ and there is only one framing variable x_1. However, the map from $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ to $H^1(\mathbf{Q}_p, k)$ is now surjective with kernel $H^1(\mathbf{Q}_p, k(\overline{\varepsilon}\lambda_{\gamma^2}))$, which is of dimension one. Note that $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = 0$, and so $R^{B_2,\urcorner}$ is formally smooth formally smooth over $\mathcal{O}$ and once again all statements are clear.

If $\gamma^2 = 1$ but $\eta \neq 0$, then $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = 0$ and the tangent space has dimension four, exactly two dimensions coming from framing, one dimension from the image of $H^1(\mathbf{Q}_p, \overline{\varepsilon})$ in $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$, and one dimension coming from the image of $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ in $H^1(\mathbf{Q}_p, k)$. To compute the image, it suffices to consider the (surjective) map from $H^1(\mathbf{Q}_p, k)$ to $H^2(\mathbf{Q}_p, \overline{\varepsilon})$ and determine the kernel, or, taking duals, considering the map $H^0(\mathbf{Q}_p, k) \to H^1(\mathbf{Q}_p, \overline{\varepsilon})$ and taking the image. The image of the latter map is precisely given by η.

The corresponding ring will fail to be flat over the space of weights Λ precisely when the image of

$$H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, k)$$

maps to zero in $H^1(I_{\mathbf{Q}_p}, k)$, or equivalently when the image is unramified. Under Tate local duality for $H^1(\mathbf{Q}_p, k) \times H^1(\mathbf{Q}_p, \overline{\varepsilon}) \to k$, the unramified classes are exactly annihilated by the peu ramifiée classes. Hence the failure of formal smoothness over Λ occurs precisely when η is peu ramifiée. All the claims follow except possibly the claim that $R^{B_2,\triangledown}$ is flat over Λ, which is also transparent except in the peu ramifée case, where $R^{B_2,\triangledown}$ is formally smooth over $\mathcal{O}$ and is the quotient of a formally smooth Λ-algebra by the relation $g_\eta = y_2 \mod (\mathfrak{m}^2, \lambda)$. This will be flat over Λ as long as $y_2 \notin \lambda R^{B_2,\triangledown}$, which can be easily ruled out by looking at points in characteristic zero. For example, we see from [GHLS17, Theorem 2.1.8] that the fibre over every point in $\Lambda[1/p]$ is non-trivial. In particular, there are points where the restriction to inertia of the character lifting λ_γ is finite of arbitrarily large order, so that $v(y_2)$ becomes arbitrarily close to 0, which would not be possible if $y_2 \in \lambda R^{B_2,\triangledown}$.

It remains to consider the case when $\overline{r} = 1 \oplus \varepsilon^{-1}$. The representation underlying $\overline{r}$ decomposes as a direct sum which induces corresponding decompositions of $\mathrm{ad}^0_{B_2}$ and $\mathrm{ad}^0_{\mathrm{GL}_2}$ respectively. In particular, the adjoint $\mathrm{ad}^0_{B_2} = k \oplus \overline{\varepsilon}$ of $\overline{r}$ thought of as inside the Borel is naturally a direct summand of $\mathrm{ad}^0_{\mathrm{GL}_2} = k \oplus \overline{\varepsilon} \oplus \overline{\varepsilon}^{-1}$. Let $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ denote the 1-cocycles with values in $\mathrm{ad}^0_{B_2}$. There is a natural surjection

$$Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = H^1(\mathbf{Q}_p, k) \oplus H^1(\mathbf{Q}_p, \overline{\varepsilon}).$$

We now chose a basis for $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ as follows:

(1) r_1 generates the kernel $B^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ of $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) \to H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$.
(2) s_1 and s_2 generate $H^1(\mathbf{Q}_p, k)$, where s_1 is unramified and s_2 is ramified.
(3) t_1 and t_2 generate $H^1(\mathbf{Q}_p, \overline{\varepsilon})$, where t_1 is très ramifiée and t_2 is peu ramifiée.
(4) Under the alternating cup product pairing

$$H^1(\mathbf{Q}_p, k) \times H^1(\mathbf{Q}_p, \overline{\varepsilon}) \to H^2(\mathbf{Q}_p, \overline{\varepsilon}) \simeq k,$$

we have $s_i \cup t_j = \delta_{ij}$.

We now define the dual basis of

$$\mathfrak{m}/(\mathfrak{m}^2, \lambda) = \mathrm{Hom}_k(\mathrm{Hom}_k(\mathfrak{m}/\mathfrak{m}^2, k), k)$$

to be given by x_i, y_i, and z_i for $i = 1$ for x_i and $i = 1, 2$ for y_i and z_i, where

$$x_i(r_j) = y_i(s_j) = z_i(t_j) = \delta_{ij},$$

and all other combinations vanish. The representation $\mathrm{ad}^0_{B_2}(\overline{\rho})$ has a Lie algebra structure via the map

$$\mathrm{ad}^0_{B_2}(\overline{\rho}) \times \mathrm{ad}^0_{B_2}(\overline{\rho}) \to \mathrm{ad}^0_{B_2}(\overline{\rho}), \qquad (A, B) \mapsto AB - BA.$$

The corresponding cup product on cohomology groups composed with this Lie algebra structure induces a symmetric bilinear pairing (the bracket cup product)

$$M : Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})^2 \to H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})^2 \to H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}),$$

Writing $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = H^1(\mathbf{Q}_p, k) \oplus H^1(\mathbf{Q}_p, \overline{\varepsilon})$, this map can be given explicitly in our case as follows:

$$M(a, b) = M((a_1, a_2), (b_1, b_2)) = 2(a_1 \cup b_2 - a_2 \cup b_1).$$

(Note that $\cup$ is alternating so this map is indeed symmetric.) As noted by in [Maz89, §1.6], the image of the corresponding map gives the quadratic relations in the deformation ring, which produces the desired quadratic relation g.

More precisely, note that $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}) = H^2(\mathbf{Q}_p, \overline{\varepsilon})$ is 1-dimensional. By [BJ15, Lem. 5.2], the relation g can be determined (up to the required order) by the relation given by the image of the natural map

$$H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})^\vee \to (Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})^\vee)^2$$

induced by the bracket cup product. But now the non-zero terms can be read off from the explicit form of $M(a, b)$ above and the description of our basis of $\mathfrak{m}/(\mathfrak{m}^2, \lambda)$ as a dual basis to the explicit basis of $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$. It follows that, after rescaling, the leading term of g is given by $y_1 z_1 + y_2 z_2$, as required.

Part (4b) is a straightforward consequences of the presentation just determined above. Note that the structure over R^{GL_1} and Λ is one again determined by the corresponding map from $H^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2})$ to $H^1(\mathbf{Q}_p, k)$, and from our explicit description above this corresponds to our choice of the parameters y_1 and y_2. □

We also have:

Lemma 7.3.9. *The points of $R^{B_2}[1/p]$ which are non-smooth over Λ are — up to unramified twist — crystalline extensions of ε^{-1} by 1.*

Proof. This is the characteristic zero version of the computation done in the proof of Lemma 7.3.7(3b), and amounts to noting that in the Tate duality pairing

$$H^1(\mathbf{Q}_p, \mathbf{Q}_p) \times H^1(\mathbf{Q}_p, \mathbf{Q}_p(1)) \to \mathbf{Q}_p,$$

the unramified classes in the first group are annihilated exactly by the crystalline extensions in the second. □

We next introduce a class of partially framed deformation rings, which will allow us to relate framed deformation rings for different groups.

7.3.10. *Partially framed deformation rings.* Since we are assuming that $F_v = \mathbf{Q}_p$, for the rest of this section we write $\mathbf{Q}_p$ instead of F_v and $\overline{\rho}$ instead of $\overline{\rho}|_{G_{F_v}}$. We shall also henceforth (in this section) write R^B for $R_v^{B,\overline{\tilde{\rho}}_v}$ and R^P for R_v^P. These are framed deformation rings with respect to GSp_4, and as such, could also be denoted by $R^{P,\square}$ and $R^{P,\square}$ to emphasize the framing. However, the images of the corresponding Galois representations may always be conjugated to land in B or P respectively. In particular, we may consider deformation rings in which the image is required to actually land inside these subgroups rather than land there up to conjugation.

Definition 7.3.11. Let $\mathcal{D}^{B,\triangleleft}$ and $\mathcal{D}^{P,\triangleleft}$ denote the subfunctors consisting of deformations which land inside B or P respectively. Let $R^{B,\triangleleft}$ and $R^{P,\triangleleft}$ denote the corresponding deformation rings.

(Here the adornment $\triangleleft$ represents that the framing is all taking place inside the "upper right corner" corresponding to B or P respectively.) The ring $R^{B,\triangleleft}$ may be identified with the universal framed deformation of $\overline{\rho}$ with fixed similitude character thought of as a *representation to* B. The ring $R^{P,\triangleleft}$ is not quite the universal deformation ring of $\overline{\rho}$ to P (framed in P) with fixed similitude character, because we are imposing an extra condition on the restriction of the first two diagonal entries to inertia. On the other hand, if $R^{P,\mathrm{univ}} = R^{P,\mathrm{univ},\square}$ denotes the deformations to GSp_4 of fixed similitude character which may be conjugated to P (without imposing this condition on inertia), then there is also a corresponding ring $R^{P,\mathrm{univ},\triangleleft}$ which is the universal P-framed deformation of $\overline{\rho}$ with fixed similitude character. Note that there are tautological maps $R^B \to R^{B,\triangleleft}$ and $R^P \to R^{P,\triangleleft}$ respectively. Let us write $\mathrm{ad}_{\mathrm{GSp}_4}$, ad_B, and ad_P for the groups $\mathrm{ad}_{\mathrm{GSp}_4}(\overline{\rho})$, $\mathrm{ad}_B(\overline{\rho})$, and $\mathrm{ad}_P(\overline{\rho})$ respectively. Since p is odd, there exist corresponding direct factors $\mathrm{ad}^0_{\mathrm{GSp}_4}$, ad^0_B, and ad^0_P corresponding to deformations with fixed similitude character.

Lemma 7.3.12. *Suppose that $\overline{\rho}$ is p-distinguished weight 2 ordinary.*

(1) There exists a splitting

$$R^{B,\triangleleft} \to R^{B,\square} \to R^{B,\triangleleft}$$

which realizes $R^B = R^{B,\square}$ as formally smooth over $R^{B,\triangleleft}$ of relative dimension

$$\dim \mathrm{ad}^0_{\mathrm{GSp}_4} - \dim \mathrm{ad}^0_B = 10 - 6 = 4.$$

(2) There exists splittings

$$R^{P,\mathrm{univ},\triangleleft} \to R^{P,\mathrm{univ},\square} \to R^{P,\mathrm{univ},\triangleleft},$$

$$R^{P,\triangleleft} \to R^{P,\square} \to R^{P,\triangleleft}$$

which realize R^P and $R^{P,\mathrm{univ}}$ as formally smooth over $R^{P,\triangleleft}$ and $R^{P,\mathrm{univ},\triangleleft}$ respectively, of relative dimension

$$\dim \mathrm{ad}^0_{\mathrm{GSp}_4} - \dim \mathrm{ad}^0_P = 10 - 5 = 5.$$

Proof. As previously noted, the p-distinguished hypothesis implies the existence (by [CHT08, Lem. 2.4.6]) of a *unique* Galois stable filtration Fil^i on $(R^{B,\square})^4$. In particular, we may choose a splitting of this filtration by a symplectic matrix $M \in \mathrm{GSp}_4(R^{B,\square})$ with the property that $M \equiv I \mod \mathfrak{m}$. Conjugation by M induces the desired map from GSp_4-framed deformations to B-framed deformations, and thus induces a splitting from $R^{B,\triangleleft}$ to R^B. In the P case, one can additionally choose the splitting such that the choice of new vector in Fil^2 is Galois stable, and then the corresponding conjugate is valued in P.

The p-distinguished hypothesis implies that the maps

$$H^0(\mathbf{Q}_p, \mathrm{ad}^0_P) \to H^0(\mathbf{Q}_p, \mathrm{ad}^0_B) \to H^0(\mathbf{Q}_p, \mathrm{ad}^0_{\mathrm{GSp}_4})$$

are all isomorphisms (see for example the explicit descriptions of ad^0_P and ad^0_B following the proof of this lemma). By construction, the reduced tangent spaces of $R^{B,\square}$ and $R^{B,\triangleleft}$ are given by extensions of $H^1(\mathbf{Q}_p, \mathrm{ad}^0_B)$ (in both cases) by $B^1(\mathbf{Q}_p, \mathrm{ad}^0_{\mathrm{GSp}_4})$

and $B^1(\mathbf{Q}_p, \mathrm{ad}^0_B)$ respectively (and analogously with P). On the other hand, the map on B^1 groups is precisely dual to the map

$$\mathrm{ad}^0_B / H^0(\mathbf{Q}_p, \mathrm{ad}^0_B) \to \mathrm{ad}^0_{\mathrm{GSp}_4} / H^0(\mathbf{Q}_p, \mathrm{ad}^0_{\mathrm{GSp}_4})$$

(and once more similarly with P). Hence, from the identification of H^0 groups above, it follows that the map on reduced tangent spaces corresponding to $R^{B,\square} \to R^{B,\daleth}$ is an injection whose cokernel has dimension

$$(\dim \mathrm{ad}^0_{\mathrm{GSp}_4} - \dim H^0(\mathbf{Q}_p, \mathrm{ad}^0_{\mathrm{GSp}_4})) - (\dim \mathrm{ad}^0_B - \dim H^0(\mathbf{Q}_p, \mathrm{ad}^0_B))$$

$$= \dim \mathrm{ad}^0_{\mathrm{GSp}_4} - \dim \mathrm{ad}^0_B$$

(And similarly in the P case with B replaced by P.)

We now prove the maps are formally smooth, which will be a direct consequence of the fact that the obstruction group is given (for $R^{B,\square}$ and $R^{B,\daleth}$ or for $R^{P,\mathrm{univ},\square}$ and $R^{P,\mathrm{univ},\daleth}$ and $R^{P,\square}$ and $R^{P,\daleth}$) by the groups $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B)$ and $H^2(\mathbf{Q}_p, \mathrm{ad}^0_P)$ respectively. We consider first the case of B; for simplicity of notation, we drop B from the superscripts from now on. Consider a surjection $\widetilde{R} := R^{\daleth}[[x_1, x_2, x_3, x_4]] \to R^{\square}$ which induces an isomorphism on reduced tangent spaces, and let J denote the kernel (so it suffices to show that $J = 0$). Let $\widetilde{\mathfrak{m}}$ be the radical of $\widetilde{R}$. Recall that we have a unique symplectic filtration Fil^i on $(R^{\square})^4 = (\widetilde{R}/J)^4$ and a choice of splitting corresponding to the matrix M. Lift this to a filtration $\widetilde{\mathrm{Fil}}^i$ for $\widetilde{R}/\widetilde{\mathfrak{m}}J$, and consider a corresponding set theoretic deformation $\widetilde{\rho} : G_{\mathbf{Q}_p} \to \mathrm{GSp}_4(\widetilde{R}/\widetilde{\mathfrak{m}}J)$ which preserves this filtration. (There are no issues lifting filtrations because the symplectic group is formally smooth.) The corresponding 2-cocycle $[c] \in H^2(\mathbf{Q}_p, \mathrm{ad}^0_{\mathrm{GSp}_4}) \otimes J/\widetilde{\mathfrak{m}}J$ then lands in $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B) \otimes J/\widetilde{\mathfrak{m}}J$. Now choose a symplectic splitting of this filtration lifting the one for Fil^i. Conjugating $\widetilde{\rho}$ by the corresponding matrix $\widetilde{M}$ (lifting M above) gives a set theoretic map $\widetilde{M}\widetilde{\rho}\widetilde{M}^{-1}$ from $G_{\mathbf{Q}_p}$ to $B(\widetilde{R}/\widetilde{\mathfrak{m}}J)$. But this map lifts $\rho^{\daleth,\mathrm{univ}} : G_{\mathbf{Q}_p} \to B(R^{\daleth})$. By universality of $\rho^{\daleth,\mathrm{univ}}$, since $\widetilde{R}/\widetilde{\mathfrak{m}}J$ is an $R^{\daleth}$-algebra, there is no obstruction to lifting this to a B-representation of $G_{\mathbf{Q}_p}$, and hence the class $[c]$ becomes trivial in the corresponding obstruction group for the B-deformation problem. Since the obstruction group in this case is $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B)$ for both the B-deformation problem and the ordinary GSp_4-deformation problem, it follows that $[c]$ is trivial and hence that $J = 0$.

The same argument applies to P, except now the splitting of $\widetilde{\mathrm{Fil}}^i$ has to be chosen so that it is preserved by $G_{\mathbf{Q}_p}$ — equivalently, an identification of the first two eigenspaces to $\widetilde{R}/J$ to ensure that the deformation is of P-type. In the case of R^P, one additionally requires the set theoretic lift to act diagonally after restriction to inertia on $\widetilde{\mathrm{Fil}}^2$. $\square$

7.3.13. *The* GSp_4*-deformation rings.* We are assuming that $\overline{\rho}|_{G_{F_v}}$ has image of the form

$$\begin{pmatrix} \lambda_{\overline{\alpha}} & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1}\eta_{\alpha\beta} & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1}\eta_{\alpha^2} \\ 0 & \lambda_{\overline{\beta}} & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1}\eta_{\beta^2} & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1}\eta_{\alpha\beta} \\ 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1} & 0 \\ 0 & 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1} \end{pmatrix}$$

where $\overline{\alpha} \neq \overline{\beta}$, and where we write η_δ to denote a (possibly zero) class in $H^1(\mathbf{Q}_p, \overline{\varepsilon}\lambda_{\overline{\delta}})$. Our analysis of the deformation rings (particularly in the B case) will depend on which of these classes are equal to zero or not.

The dimensions of B and P are 7 and 6 respectively. Recall that we are considering deformations of $\overline{\rho}$ to B or P with fixed similitude character. For $p > 2$, the adjoint representations $\mathfrak{p}$ and $\mathfrak{b}$ admit a splitting with a canonical one dimensional summand corresponding to varying the similitude character. Let $\mathrm{ad}^0_B(\overline{\rho}) \subset \mathfrak{b}$ and $\mathrm{ad}^0_P(\overline{\rho}) \subset \mathfrak{p}$ denote the complementary 6 and 5 dimensional subspaces. Explicitly, $\mathfrak{b}$ is given as follows:

$$\mathfrak{b} = \begin{pmatrix} \nu + x_\alpha & -x_{\alpha/\beta} & x_{\alpha\beta} & x_{\alpha^2} \\ 0 & \nu + x_\beta & x_{\beta^2} & x_{\alpha\beta} \\ 0 & 0 & \nu - x_\beta & x_{\alpha/\beta} \\ 0 & 0 & 0 & \nu - x_\alpha \end{pmatrix},$$

where the subspace with $x_{\alpha/\beta} = 0$ corresponds to $\mathfrak{p}$, and the subspace $\nu = 0$ corresponds to ad^0_B. With respect to the basis given by the matrices corresponding to $\{x_{\alpha^2}, x_{\beta^2}, x_{\alpha\beta}, x_{\alpha/\beta}, x_\alpha, x_\beta\}$, the Galois representation $\mathrm{ad}^0_B(\overline{\rho})$ is given explicitly as follows:

$$\begin{pmatrix} \overline{\varepsilon}\lambda_{\overline{\alpha}^2} & 0 & 0 & 2\lambda_{\overline{\alpha}}\lambda_{\overline{\beta}}^{-1} \cdot \eta_{\alpha\beta} & -2\eta_{\alpha^2} & 0 \\ 0 & \overline{\varepsilon}\lambda_{\overline{\beta}^2} & 0 & 0 & 0 & -2\eta_{\beta^2} \\ 0 & 0 & \overline{\varepsilon}\lambda_{\overline{\alpha}\overline{\beta}} & \lambda_{\overline{\alpha}}\lambda_{\overline{\beta}}^{-1} \cdot \eta_{\beta^2} & -\eta_{\alpha\beta} & -\eta_{\alpha\beta} \\ 0 & 0 & 0 & \lambda_{\overline{\alpha}}\lambda_{\overline{\beta}}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

and, on the space $\mathrm{ad}^0_P(\overline{\rho}) \subset \mathrm{ad}^0_B(\overline{\rho})$ with respect to the basis $\{x_{\alpha^2}, x_{\beta^2}, x_{\alpha\beta}, x_\alpha, x_\beta\}$ (not a direct summand!), we have

$$\begin{pmatrix} \overline{\varepsilon}\lambda_{\overline{\alpha}^2} & 0 & 0 & -2\eta_{\alpha^2} & 0 \\ 0 & \overline{\varepsilon}\lambda_{\overline{\beta}^2} & 0 & 0 & -2\eta_{\beta^2} \\ 0 & 0 & \overline{\varepsilon}\lambda_{\overline{\alpha}\overline{\beta}} & -\eta_{\alpha\beta} & -\eta_{\alpha\beta} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

As in the case of GL_2 above, we may compute $H^2(\mathbf{Q}_p, \mathrm{ad}^0(\overline{\rho}))$ for P and B by counting whether the subspaces generated by $\{x_{\alpha^2}, x_{\beta^2}, x_{\alpha\beta}\}$ generate $\overline{\varepsilon}$ subspaces and whether these subspaces split. The following lemma is immediate from the explicit description above.

Lemma 7.3.14. *The dimension of $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(\overline{\rho}))$ is zero unless one of the following holds:*

(1) The classes $\eta_{\alpha\beta}$ and η_{α^2} are both zero, and $\overline{\alpha}^2 = 1$.

(2) The class η_{β^2} is zero, and $\overline{\beta}^2 = 1$. In this case, either:

(a) The conditions of part (1) also hold, or:

(b) The dimension of $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(\overline{\rho}))$ is 1, and there is a $\mathbf{Q}_p$-equivariant map from the representation V underlying $\overline{\rho}$ to the Borel of $\mathrm{GL}(2)$ corresponding to the representation

$$W = \begin{pmatrix} \lambda_{\overline{\beta}} & 0 \\ 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1} \end{pmatrix} = \lambda_{\overline{\beta}} \otimes \begin{pmatrix} 1 & 0 \\ 0 & \overline{\varepsilon}^{-1} \end{pmatrix},$$

and the corresponding map relating $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(\overline{\rho})) = H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(V))$ *to* $H^2(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}(W))$ *is an isomorphism.*

(3) The classes $\eta_{\alpha\beta}$ *and* η_{β^2} *are both zero, and* $\overline{\alpha}\overline{\beta} = 1$.

Moreover, the dimension of $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(\overline{\rho}))$ *is* ≥ 2 *only in case (2a), in which it has dimension* 2.

We could give a similar (but easier) computation of $H^2(\mathbf{Q}_p, \mathrm{ad}^0_P(\overline{\rho}))$, but it is not needed in the sequel so it is omitted.

Recall that $R^{P,\mathrm{univ}}$ denotes the universal deformation ring for P. The quotient R^P is given by imposing the condition that the action of inertia on Fil^2 (given by the upper left 2×2 matrix after changing basis) is through a scalar. Recall that we also have corresponding rings $R^{P,\mathrm{univ},\triangleleft}$ and $R^{P,\triangleleft}$ where the image lands in P directly (rather than up to conjugation). Any deformation of type $R^{P,\mathrm{univ},\triangleleft}$ determines deformations of the three 2-dimensional subquotients of $\overline{\rho}$, given respectively by the extension $\overline{r}_A$ of $\overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}}^{-1}$ by $\lambda_{\overline{\alpha}}$, by the extension $\overline{r}_B$ of $\overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1}$ by $\lambda_{\overline{\beta}}$, and the extension $\overline{r}_{AB}$ of $\overline{\varepsilon}^{-1}\lambda_{\overline{\beta}}^{-1}$ by $\lambda_{\overline{\alpha}}$. Similarly, any triple of such deformations with the appropriate coincidences of the corresponding characters defines a representation of type P. (These identifications require that we work with $\triangleleft$ framings rather than $\square$ framings, since otherwise there would be superfluous framing variables in this identification.)

Let $R_A = R^{B_2,\triangleleft}$ for $\overline{r}_A$ and $R_B = R^{B_2,\triangleleft}$ for $\overline{r}_B$. Let $R_{AB} = R^{B_2,\triangleleft,\det}$ for $\overline{r}_{AB}$, where $R^{B_2,\triangleleft,\det}$ is the framed B_2 deformation ring in which one does not fix the determinant, so (since $p > 2$) one has that $R^{B_2,\triangleleft,\det} = R^{B_2,\triangleleft}\widehat{\otimes}_{\mathcal{O}} R^{\mathrm{GL}_1}$ is formally smooth over $R^{B_2,\triangleleft}$ of relative dimension 2. There are natural maps from R_A, R_B, and R_{AB} to $R^{P,\mathrm{univ},\triangleleft}$ and $R^{P,\triangleleft}$ respectively. Write $R^{\mathrm{GL}_1 \times \mathrm{GL}_1} = R^{\mathrm{GL}_1}\widehat{\otimes}_{\mathcal{O}} R^{\mathrm{GL}_1}$ for the deformation ring corresponding to the pair of characters $(\lambda_{\alpha_v}, \lambda_{\beta_v})$. We have the following:

Lemma 7.3.15. *The ring* $R^{P,\mathrm{univ}}$ *is formally smooth over*

$$R^{P,\mathrm{univ},\triangleleft} \simeq (R_A \widehat{\otimes}_{\mathcal{O}} R_B)\widehat{\otimes}_{R^{\mathrm{GL}_1 \times \mathrm{GL}_1}} R_{AB}$$

of relative dimension 5. *The ring* R^P *is formally smooth over*

$$R^{P,\triangleleft} \simeq (R_A \widehat{\otimes}_{\Lambda} R_B)\widehat{\otimes}_{R^{\mathrm{GL}_1 \times \mathrm{GL}_1}} R_{AB}$$

of relative dimension 5.

Proof. The isomorphisms follow directly from the discussion above, and the statement about formal smoothness is Lemma 7.3.12. $\square$

We now prove the P-part of Proposition 7.3.4.

Proposition 7.3.16. *Suppose that* $\overline{\rho}$ *is* p*-distinguished. Then* $R^{P,\mathrm{univ}}$ *and* R^P *are both complete intersections. Moreover, they are connected in characteristic zero, and the non-smooth locus in characteristic zero has codimension at least two. In particular, the generic fibres* $R^{P,\mathrm{univ}}[1/p]$ *and* $R^P[1/p]$ *are irreducible of dimensions* 15 *and* 14 *over* $\mathbf{Q}_p$ *respectively.*

Proof. The strategy is as follows. By Lemma 7.3.15, we can immediately reduce to the rings $R^{P,\mathrm{univ},\triangleleft}$ and $R^{P,\triangleleft}$ respectively and prove that they satisfy the same properties above (with 15 and 14 replaced by 10 and 9 respectively). Given the explicit form of the presentations for the 2-dimensional B_2-deformation rings, in

order to show that $R^{P,\mathrm{univ},\urcorner}$ and $R^{P,\urcorner}$ are complete intersections, one can simply write down enough about the equations for the tensor products in Lemma 7.3.15 and observe (for the appropriate value $d = 10$ or 9 in either case) that they are either:

(1) Formally smooth of the relative dimension d over $\mathcal{O}$,
(2) Given as a quotient of a power series ring in $d+1$ variables by one relation,
(3) Given as a quotient of a power series ring in $d+2$ variables by a 2-generator prime ideal which is not contained in (λ).

(The last example occurs only in a single case.) We say more about this computation below.

For the remaining claims, it suffices to prove that the generic fibre is connected and that our tensor products are R_1 and S_2 (and thus normal); since they are complete intersections, it is enough to show that the non-smooth points have codimension at least 2. It is convenient to consider two separate cases.

Suppose that $\alpha\beta = 1$. In this case, it follows from the p-distinguishedness hypothesis that $\alpha^2 \neq 1$ and $\beta^2 \neq 1$. In this case, the rings above have a particularly simple form even over $\mathcal{O}$. Namely, R_A and R_B are formally smooth over $\mathcal{O}$ and over Λ, and the resulting tensor product is formally smooth over R_{AB}, and thus the result follows from Lemma 7.3.7, since the rings $R^{B_2,\urcorner}$ satisfy all the required geometric properties above.

Now suppose that $\alpha\beta \neq 1$. In this case, R_{AB} is formally smooth over $R^{\mathrm{GL}_1 \times \mathrm{GL}_1}$, and by Lemma 7.3.15, $R^{P,\mathrm{univ}}$ and $R^{P,\urcorner}$ are formally smooth over $R_A \widehat{\otimes}_{\mathcal{O}} R_B$ or $R_A \widehat{\otimes}_{\Lambda} R_B$ respectively. Let us now consider the case of $R^{P,\urcorner}$, which corresponds to $R_A \widehat{\otimes}_{\Lambda} R_B$, the case of $R_A \widehat{\otimes}_{\mathcal{O}} R_B$ being easier and also following immediately from Lemma 7.3.7. Since R_A and R_B are either smooth or have a non-smooth locus of codimension 4 (corresponding to twists of $1 \oplus \varepsilon^{-1}$ by a (possibly trivial) unramified quadratic character), it is certainly the case that the points on the generic fibre of $R_A \widehat{\otimes}_{\Lambda} R_B$ which are non-smooth on $R_A \widehat{\otimes}_{\mathcal{O}} R_B$ have codimension at least 2. Hence it suffices to consider the non-smooth points of $R_A \widehat{\otimes}_{\Lambda} R_B$ which are smooth on $R_A \widehat{\otimes}_{\mathcal{O}} R_B$. In particular, such a point must have a tangent space of dimension 8, and will be smooth if and only if it has an infinitesimal deformation which does not lie on $R_A \widehat{\otimes}_{\Lambda} R_B$. Equivalently, given a point $x = (x_A, x_B)$ on the generic fibre of $R_A \widehat{\otimes}_{\Lambda} R_B[1/p]$, it will be smooth if it has a deformation in which the weight over Λ varies for one x_A or x_B but remains fixed for the other point. Equivalently, we can look for a deformation of $x = (x_A, x_B)$ such that one point is fixed but the other point varies over $\Lambda[1/p]$. For x_A or x_B, such a deformation exists as long as x_A (or x_B) is a smooth point over Λ. But the non-smooth points in characteristic zero over the space of weights Λ are (up to unramified twist) exactly the crystalline extensions of ε^{-1} by 1 (see Lemma 7.3.9), and hence these non-smooth points certainly have codimension at least 2. To show it is connected, it suffices to note that, for each fibre of R_A above Λ, any x_A is connected over this fibre to a point which is smooth over Λ. This reduces to showing that any extension of ε^{-1} by 1 which is crystalline has a deformation to a non-crystalline extension. But this is trivially achieved by a perturbation of the extension class, noting that $H^1(\mathbf{Q}_p, \varepsilon)$ is free of rank 2 and the crystalline subspace is a line of rank 1.

It remains to prove the claim that these rings are complete intersections in all the possible cases. Almost all the time, the tensor product is either immediately seen to be formally smooth of the right dimension, or given by a single non-zero

equation and of the right dimension. In fact, the only way in which there can be two equations is when two of the rings R_A, R_B, and R_{AB} are not formally smooth. This implies that at least two of $\overline{\alpha}^2$, $\overline{\beta}^2$, and $\overline{\alpha}\overline{\beta}$ are equal to one, and this trivially only happens when $\overline{\alpha}^2 = 1$ and $\overline{\beta}^2 = 1$, and hence $\alpha\beta \neq 1$. Thus the only possible case when there exist at least two equations is when $\overline{\alpha}^2 = \overline{\beta}^2 = 1$ and $\eta_{\overline{\alpha}^2} = \eta_{\overline{\beta}^2} = 0$. The corresponding tensor product is then

$$\mathcal{O}[[x_{A,1}, z_{A,1}, z_{A,2}, y_{A,1}, y_{A,2}, x_{B,1}, z_{B,1}, z_{B,2}, y_{B,1}]]$$

modulo the ideal (noting tensoring over Λ forces $y_{A,2} = y_{B,2}$):

$$(z_{A,1}y_{A,1} + z_{A,2}y_{A,2} + \ldots, z_{B,1}y_{B,1} + z_{B,2}y_{A,2} + \ldots).$$

This pair of elements is easily seen to generate a height 2 prime ideal.

The cases when there are no equations and the rings are formally smooth are trivial. In the cases when there is an extra generator one has to show that the resulting equation is non-zero. Essentially the most subtle case of this form occurs when $\overline{\alpha}^2 = \overline{\beta}^2 = 1$ and $\eta_{\overline{\alpha}^2}$ and $\eta_{\overline{\beta}^2}$ are both non-zero and peu ramifiée. In that case, there are naïvely two equations which have the following form:

$$y_{A,2} = h(x_{A,1}, z_{A,1}, z_{A,2}, y_{A,1}),$$
$$y_{A,2} = h(x_{B,1}, z_{B,1}, z_{B,2}, y_{B,1}),$$

which immediately reduces to one equation. (Here h is the same h because both extensions generate the same line — the other cases are trivial). We then need to show that the resulting equation obtained by taking the difference of the RHS is non-zero. But this is obviously the case unless the RHS is zero. If this is true, then $y_{A,2}$ is zero in $R^{B_2,\triangleleft}/\lambda$, which is impossible since $R^{B_2,\triangleleft}$ is flat over Λ. □

Proof of Proposition 7.3.4. By Proposition 7.3.16, we only need to prove the results for R^B. As in the proof of Lemma 7.3.6, we have a presentation of R^B of the form $\mathcal{O}[[x_1, \ldots, x_r]]/(y_1, \ldots, y_s)$, where

$$r = 11 - h^0(\mathbf{Q}_p, \mathrm{ad}_B\,\overline{\rho}) + h^1(\mathbf{Q}_p, \mathrm{ad}^0_B\,\overline{\rho}),\ \ s = h^2(\mathbf{Q}_p, \mathrm{ad}^0_B\,\overline{\rho}),$$

so that by the local Euler characteristic formula, $r - s = 10 + \dim \mathrm{ad}^0_B\,\overline{r} = 16$. In particular, if $H^2(\mathbf{Q}_p, \mathrm{ad}^0_B(\overline{\rho})) = 0$, then R^B is formally smooth over $\mathcal{O}$ of relative dimension 16, and there is nothing to prove.

It is therefore enough to consider each of the cases of Lemma 7.3.14. In case (2b), we see that R^B is formally smooth of relative dimension 11 over the deformation ring $R^{B_2,\triangleleft}$ for $\overline{r} = \lambda_{\overline{\beta}} \oplus \lambda_{\overline{\beta}}^{-1}\overline{\varepsilon}^{-1}$, so the result follows from Lemma 7.3.6. From the presentation in the previous paragraph, we see that in cases (1) and (3), R^B is a complete intersection, while in case (2a), we see that every irreducible component of R^B has relative dimension at least 16 over $\mathcal{O}$. By Lemma 7.3.12, we may (and we do) pass freely between R^B and $R^{B,\triangleleft}$ when convenient.

Suppose that we are in case (3), and suppose that $\eta_{\alpha^2} \neq 0$. Let $\overline{\varrho}$ be the representation with the same $\overline{\alpha}, \overline{\beta}$ as $\overline{\rho}$, but with $\eta_{\alpha^2} = \eta_{\beta^2} = \eta_{\alpha\beta} = 0$. Let $R^{B,\triangleleft}_{\overline{\varrho}}$ be the corresponding deformation ring. We claim that in fact $R^{B,\triangleleft}_{\overline{\varrho}}$ and $R^{B,\triangleleft}$ are isomorphic. To see this, note firstly that since $\overline{\alpha}\overline{\beta} = 1$, and $\overline{\alpha} \neq \overline{\beta}$, we have $\overline{\alpha}^2 \neq 1$. Let

$$\overline{r} = \begin{pmatrix} \lambda_{\overline{\alpha}} & \eta_{\alpha^2} \\ 0 & \lambda_{\overline{\alpha}}^{-1}\overline{\varepsilon}^{-1} \end{pmatrix}.$$

We have already shown that, in this case, $R^{B_2,\triangledown}$ is formally smooth over Λ. In fact, we can be more explicit. Write $\widetilde{\Lambda}$ for what we called R^{GL_1} above, so that $\widetilde{\Lambda}$ is the formally smooth Λ-algebra of relative dimension 1 which carries the additional information of the actual lift of $\lambda_{\overline{\alpha}}$ (rather than just its restriction to inertia), so that $R^{B,\triangledown}$ and $R^{B_2,\triangledown}$ are naturally $\widetilde{\Lambda}$-algebras. Let $\widetilde{\lambda_{\overline{\alpha}}} : G_{\mathbf{Q}_p} \to \widetilde{\Lambda}^\times$ be the universal lift of $\lambda_{\overline{\alpha}}$. Then since $\overline{\alpha}^2 \neq 1$, $H^1(G_{\mathbf{Q}_p}, \widetilde{\lambda_{\overline{\alpha}}}^2 \varepsilon)$ is a free $\widetilde{\Lambda}$-module of rank 1, and the universal lift of $\overline{r}$ is represented by

$$\begin{pmatrix} \widetilde{\lambda_{\overline{\alpha}}} & \widetilde{\lambda_{\overline{\alpha}}}^{-1} \varepsilon^{-1} \widetilde{\eta_{\alpha^2}} \\ 0 & \widetilde{\lambda_{\overline{\alpha}}}^{-1} \overline{\varepsilon}^{-1} \end{pmatrix}$$

where $\widetilde{\eta_{\alpha^2}}$ lifts η_{α^2}. It is then easy to verify that if ϱ^{univ} is the universal upper-triangular lift of $\overline{\varrho}$ to $R^{B,\triangledown}_{\overline{\varrho}}$, then

$$\varrho^{\mathrm{univ}} + \begin{pmatrix} 0 & 0 & 0 & \widetilde{\lambda_{\overline{\alpha}}}^{-1} \varepsilon^{-1} \widetilde{\eta_{\alpha^2}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

is a lift of $\overline{\rho}$. This gives a map $R^{B,\triangledown}_{\overline{\varrho}} \to R^{B,\triangledown}$, and we can obtain a map $R^{B,\triangledown} \to R^{B,\triangledown}_{\overline{\varrho}}$ in the same way. It is clear that the composites of these maps are the identities, so that $R^{B,\triangledown}_{\overline{\varrho}}$ and $R^{B,\triangledown}$ are isomorphic, as claimed.

Accordingly, whenever we are in case (3), we will assume from now on that $\eta_{\alpha^2} = 0$. It is now easy to see that in each of the cases (1), (2a), and (3), $\operatorname{Spec} R^{B,\triangledown}[1/p]$ is connected. Indeed, in each case we have $\eta_{\alpha^2} = \eta_{\alpha\beta} = 0$, so by arguing as in the proof of Lemma 7.3.6 (using conjugation by $\operatorname{diag}(t, 1, 1, t^{-1})$), we see that every closed point of $\operatorname{Spec} R^{B,\triangledown}[1/p]$ may be path connected to one which lands in $P(\overline{\mathbf{Q}}_p)$. Now we may immediately conclude by knowing the corresponding result for $\operatorname{Spec} R^{P,\mathrm{univ},\triangledown}[1/p]$ proved in Proposition 7.3.16.

To obtain irreducibility we now argue as in the proof of Lemma 7.3.6, by studying the singular locus of $\operatorname{Spec} R^B[1/p]$. More precisely, we let $\rho^{\mathrm{univ}} : G_{\mathbf{Q}_p} \to \mathrm{GSp}_4(R^B)$ be the universal lifting, and let $H^2 := H^2(G_{\mathbf{Q}_p}, \operatorname{ad}^0_B \rho^{\mathrm{univ}})$, a finite R^B-module, which is compatible with specialization. Let U be the complement of the support of H^2 in $\operatorname{Spec} R^B[1/p]$. At any closed point $x \in U$ with corresponding representation $\rho_x : G_{\mathbf{Q}_p} \to \mathrm{GSp}_4(E_x)$, we have $H^2(G_{\mathbf{Q}_p}, \operatorname{ad}^0_B \rho_x) = 0$, so it follows that U is formally smooth over E_x at x of relative dimension 16. In particular, U is regular.

The points in the complement of U are those for which $H^2(G_{\mathbf{Q}_p}, \operatorname{ad}^0_B \rho_x) \neq 0$. We claim that this has codimension at least 2. We may explicitly describe this locus as follows (this description follows easily from the explicit description of ad^0_B preceding Lemma 7.3.14). In case (1), we may suppose without loss of generality (by twisting with a quadratic character if necessary) that $\overline{\alpha} = 1$, and then the points in the complement of U are those conjugate to representations of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \chi & * & 0 \\ 0 & 0 & \varepsilon^{-1}\chi^{-1} & 0 \\ 0 & 0 & 0 & \varepsilon^{-1} \end{pmatrix}$$

where χ lifts $\lambda_{\overline{\beta}}$. The locus of such points has dimension at most 13; indeed, the action of PGSp_4 by conjugation contributes at most 10 to the dimension, and the

choice of χ and $*$ at most 3 (there is a two-dimensional family of choices of χ, and if χ^2 is non-trivial then the choice of $*$ gives one more dimension, while if χ^2 is trivial then it gives 2 dimensions).

In case (2a), we have in addition the points of the form

$$\begin{pmatrix} \chi' & * & 0 & 0 \\ 0 & \lambda_{-1} & 0 & 0 \\ 0 & 0 & \varepsilon^{-1}\lambda_{-1} & * \\ 0 & 0 & 0 & \varepsilon^{-1}(\chi')^{-1} \end{pmatrix}$$

where χ' lifts $\lambda_{\overline{\alpha}} = 1$. The locus of such points again has dimension at most 13.

In case (3), we have the points of the form

$$\begin{pmatrix} \chi & * & 0 & * \\ 0 & \chi^{-1} & 0 & 0 \\ 0 & 0 & \varepsilon^{-1}\chi & * \\ 0 & 0 & 0 & \varepsilon^{-1}\chi^{-1} \end{pmatrix},$$

where χ lifts $\lambda_{\overline{\alpha}}$. The locus of such points has dimension at most 14 (with 2 dimensions for the choice of χ, and then generically one dimension each for the choices of the extension class of χ^{-1} by χ and of $\varepsilon^{-1}\chi^{-1}$ by χ, or two dimensions each if $\chi^2 = 1$).

Thus in cases (1) and (3), since we know that R^B is a complete intersection, we see that it is normal (being R_1 and S_2), so we are done. The case (2b) having already been dealt with, we are left with case (2a), where we have seen that every irreducible component of $\operatorname{Spec} R^B[1/p]$ has dimension at least 16, while the complement of U has dimension at most 12. It now suffices to show that R^B is a complete intersection, and thus also normal as above, and to check that $\operatorname{Spec} R^B[1/p]$ has dimension exactly 16.

We have a presentation of $R = R^B$ of the form $\mathcal{O}[[x_1, \ldots, x_{18}]]/(y_1, y_2)$. This is a complete intersection as long as $\dim(R) \leq 19 - 2 = 17$, which also implies that the relative dimension of R over $\mathcal{O}$ is 16, and so the dimension of the generic fibre is 16. Assume otherwise, so that $\dim(R) \geq 18$. Then the support of R in $\operatorname{Spec} \mathcal{O}[[x_1, \ldots, x_{18}]]$ contains a height one prime $\mathfrak{p}$ of $\mathcal{O}[[x_1, \ldots, x_{18}]]$. Suppose firstly that $\mathfrak{p}$ has residue characteristic zero, and let T denote the corresponding closed subscheme of $\operatorname{Spec} R[1/p]$, which will have dimension 17. For any closed point $x \in T$ with corresponding representation $\rho_x : G_{\mathbf{Q}_p} \to \mathrm{GSp}_4(E_x)$, the tangent space at x certainly has dimension at least $\dim(T) = 17$. Hence there is an inequality

$$11 - h^0(\mathbf{Q}_p, \operatorname{ad}_B \rho_x) + h^1(\mathbf{Q}_p, \operatorname{ad}^0_B \rho_x) \geq 17,$$

and so, by the Euler characteristic formula, $h^2(G_{\mathbf{Q}_p}, \operatorname{ad}^0_B \rho_x) \geq 17 - 16 \geq 1$. In particular, it follows that x lies in the support of H^2, and hence that $T \subset U$. But we have already seen that U has dimension at most 12, and this is a contradiction.

Hence R can only fail to be a complete intersection if the support of R contains (λ). It follows that $\dim(R/\lambda) = \dim(k[[x_1, \ldots, x_{18}]])$, and hence that $R/\lambda = k[[x_1, \ldots, x_{18}]]$. Twisting, we may without loss of generality assume that $\overline{\beta} = 1$. Let $\overline{r} = 1 \oplus \overline{\varepsilon}^{-1}$, and let $R^{B_2, \urcorner}$ denote the corresponding fixed determinant deformation ring to the Borel of GL(2). By realizing $\overline{r}$ as the subquotient of the representation $\overline{\rho}$ given by the span of the second and third standard basis vectors, there is an induced map

$$\psi : R^{B_2, \urcorner} \to R^{\urcorner} \to R \to R/\lambda = k[[x_1, \ldots, x_{18}]].$$

Let W denote the representation underlying $\overline{r}$, and (as previously) V the representation underlying $\overline{\rho}$. Let us now consider the induced map on reduced tangent spaces. To compute this, we may look at the corresponding deformation rings, and consider the induced map on tangent spaces. For $R^{B_2,\triangledown}$, the tangent space is given by $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}(W))$. For $R^{\triangledown}$, it is given by $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_B(V))$. Note that we are assuming that $\eta_{\alpha\beta} = \eta_{\alpha^2} = \eta_{\beta^2} = 0$, and so $\overline{\rho}$ is completely split, and so $\mathrm{ad}^0_{B_2}$ is a direct summand of ad^0_B. Thus $Z^1(\mathbf{Q}_p, \mathrm{ad}^0_{B_2}(W)) \to Z^1(\mathbf{Q}_p, \mathrm{ad}^0_B(V))$ is injective. On the other hand, Lemma 7.3.7 (4b) shows that (for this $\overline{r}$) the ring $R^{B_2,\triangledown}/\lambda$ is *not* formally smooth. But this is a contradiction; a minimal set of generators of the maximal ideal of $R^{B_2,\triangledown}/\lambda$ satisfy at least one polynomial relation, but their images under ψ do not satisfy any such relation under our assumptions because the map on tangent spaces is injective and (as we are currently assuming) $R^{\triangledown}/\lambda$ and R/λ is formally smooth. Hence λ also cannot be in the support of R, and thus R is a complete intersection. □

Remark 7.3.17. The last argument shows that, in case (2a), the ring $R = R^B$ is a complete intersection. But we certainly expect (in this and in all other cases) the stronger properties that R is flat over $\mathcal{O}$ and R/λ is also a complete intersection, whereas the argument only shows that $\dim(R/\lambda) \leq 17$, rather than $\dim(R)-1 = 16$, which would be necessary in order for λ to be a regular element. In general, we have often only attempted to prove exactly enough about the deformation rings that we require for the argument, rather than giving a fuller account of their geometric properties. We apologize to readers who examine this argument in closer detail who were hoping for something more comprehensive.

As in §7.1, we say that a closed point x of $R_v^{B,\bar{\tilde{r}}_v}[1/p]$ (resp. $R_v^P[1/p]$) is smooth if $(R_v^{B,\bar{\tilde{r}}_v}[1/p])_x$ is regular (resp. $(R_v^{B,\bar{\tilde{r}}_v}[1/p])_x$ is regular). We say that the corresponding Galois representation ρ_x is pure if it is de Rham, and if $\mathrm{WD}(\rho_x)$ is pure (that is, it arises as the base extension of a pure Weil–Deligne representation over a number field).

Lemma 7.3.18. *If x is a closed point of the generic fibre of $R_v^{B,\bar{\tilde{r}}_v}[1/p]$ or $R_v^P[1/p]$, and x is pure, then it is smooth.*

Proof. We first consider the case of B. From the proof of Proposition 7.3.4, we see that it is enough to check that $H^2(G_{\mathbf{Q}_p}, \mathrm{ad}^0_B \rho_x) = 0$. By Tate local duality, this means that it is enough to check that $\mathrm{Hom}_{G_{\mathbf{Q}_p}}(\rho_x, \rho_x(1)) = 0$, and therefore it is enough to check that $\mathrm{Hom}_{\mathrm{WD}_{\mathbf{Q}_p}}(\mathrm{WD}(\rho_x), \mathrm{WD}(\rho_x(1))) = 0$. This follows easily from the definition of purity. The same argument also applies to $R_v^{P,\mathrm{univ}}[1/p]$. We now consider $R_v^P[1/p]$. The non-smooth points x of $R_v^P[1/p]$ are either non-smooth in $R_v^{P,\mathrm{univ}}[1/p]$ (for which the previous argument applies) or, via the isomorphism of Lemma 7.3.15 and the proof of Proposition 7.3.16, arise in the following way: the representation ρ_x admits a 2-dimensional reducible subquotient r_x such that the corresponding point on the deformation ring $R^{B_2}[1/p]$ is not smooth over Λ_v. By Lemma 7.3.9, such representations are (up to unramified twist) a crystalline extension of ε^{-1} by 1. Since these are not pure (and purity is preserved by taking subquotients), the representation ρ_x is also not pure, and we are also done in this case. □

7.4. **Local deformation problems, $l \neq p$.**

7.4.1. *Unobstructed deformations.* Assume that $v \nmid p$.

Proposition 7.4.2. *If $H^0(F_v, \mathrm{ad}^0\,\overline{\rho}(1)) = 0$, then $R_v^{\square}$ is isomorphic to a power series ring over $\mathcal{O}$ in 10 variables. If furthermore $\overline{\rho}|_{G_{F_v}}$ is unramified, then so are all of its lifts.*

Proof. By Tate duality, the condition is equivalent to $H^2(F_v, \mathrm{ad}^0\,\overline{\rho}) = 0$, and the result follows from a standard calculation in obstruction theory (see e.g. [Til96, §5.2]). □

7.4.3. *Taylor–Wiles deformations.* Assume that $q_v \equiv 1 \bmod p$, and that both $\psi|_{G_{F_v}}$ and $\overline{\rho}|_{G_{F_v}}$ are unramified. We take $\Lambda_v = \mathcal{O}$. We assume that $\overline{\rho}(\mathrm{Frob}_v)$ has 4 distinct eigenvalues in k, and we fix an ordering of them as $\overline{\alpha}_1$, $\overline{\alpha}_2$, $\overline{\alpha}_3 = \psi(\mathrm{Frob}_v)/\overline{\alpha}_2$, $\overline{\alpha}_4 = \psi(\mathrm{Frob}_v)/\overline{\alpha}_1$. For each $i = 1, 2$, let $\overline{\gamma}_i : G_{F_v} \to k^\times$ be the unramified character that sends Frob_v to $\overline{\alpha}_i$.

Lemma 7.4.4. *Let $\rho : G_{F_v} \to \mathrm{GSp}_4(A)$ be any lift of $\overline{\rho}$. There are unique continuous characters $\gamma_i : G_{F_v} \to A^\times$ for $i = 1, 2$, such that ρ is $\mathrm{GSp}_4(A)$-conjugate to a lift of the form $\gamma_1 \oplus \gamma_2 \oplus \psi\gamma_2^{-1} \oplus \psi\gamma_1^{-1}$, where $\gamma_i \bmod \mathfrak{m}_A = \overline{\gamma}_i$ for each $i = 1, 2$.*

Proof. This can be proved in exactly the same way as [GT05, Lem. 5.1.1]. □

Let $\Delta_v = k(v)^\times(p)^2$, where $k(v)^\times(p)$ is the maximal p-power quotient of $k(v)^\times$, and let $\rho : G_{F_v} \to \mathrm{GSp}_4(R_v^{\square})$ denote the universal lift. Then ρ is $\mathrm{GSp}_4(R_v^{\square})$-conjugate to a lift of the form $\gamma_1 \oplus \gamma_2 \oplus \psi\gamma_2^{-1} \oplus \psi\gamma_1^{-1}$ as in Lemma 7.4.4. For $i = 1$, 2, the character $\gamma_i \circ \mathrm{Art}_{F_v}|_{\mathcal{O}_{F_v}^\times}$ factors through $k(v)^\times(p)$, so we obtain a canonical local $\mathcal{O}$-algebra morphism $\mathcal{O}[\Delta_v] \to R_v^{\square}$. Note that this depends on the choice of ordering $\overline{\alpha}_1, \ldots, \overline{\alpha}_4$. It is straightforward to check that this morphism is formally smooth of relative dimension 10.

7.4.5. *Ihara avoidance deformations.* Let v be a finite place of F with $q_v \equiv 1 \bmod p$. Assume further that $\overline{\rho}|_{G_{F_v}}$ is trivial, and that $\psi|_{G_{F_v}}$ is unramified and has trivial reduction modulo λ. We take $\Lambda_v = \mathcal{O}$.

Let $\chi = (\chi_1, \chi_2)$ be a pair of continuous characters $\chi_i : \mathcal{O}_{F_v}^\times \to \mathcal{O}^\times$ that are trivial modulo λ. We let $\mathcal{D}_v^\chi$ be the functor of lifts $\rho : G_{F_v} \to \mathrm{GSp}_4(A)$ such that for all $\sigma \in I_{F_v}$, the characteristic polynomial of $\rho(\sigma)$ is

$$(X - \chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1})(X - \chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}).$$

Then $\mathcal{D}_v^\chi$ is a local deformation problem, and we denote its representing object by R_v^χ.

Lemma 7.4.6. *If $\chi_1, \chi_2 \neq 1$ and $\chi_1 \neq \chi_2^{\pm 1}$, then every closed point of $\mathrm{Spec}\, R_v^\chi[1/p]$ is smooth.*

Proof. We can choose $\sigma \in I_{F_v}$ with $\chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma))$, $\chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma))$, $\chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}$, $\chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}$ pairwise distinct. As in Lemma 7.1.3, we need to check that for every point x, we have $\mathrm{Hom}_{E'[G_{F_v}]}(\rho_x, \rho_x(1)) = 0$. Any such homomorphism would have to respect the eigenspaces for $\rho_x(\sigma)$, and must therefore be zero. □

The proof of the following two results occupies the rest of this subsection.

Proposition 7.4.7. *Assume that $\chi_1 = \chi_2 = 1$. Then R_v^1 satisfies the following properties:*

(1) $\operatorname{Spec} R_v^1$ *is equidimensional of dimension* 11 *and every generic point has characteristic zero.*
(2) Every generic point of $\operatorname{Spec} R_v^1/(\lambda)$ *is the specialization of a unique generic point of* $\operatorname{Spec} R_v^1$.

Proposition 7.4.8. *Assume that* $\chi_1, \chi_2 \neq 1$ *and* $\chi_1 \neq \chi_2^{\pm 1}$. *Then* $\operatorname{Spec} R_v^\chi$ *is irreducible of dimension* 11*, and its generic point has characteristic zero.*

We follow the strategy of [Tay08] (which proves the corresponding results for GL_n) closely. A source of minor complications in the case of GSp_4 is that nilpotent centralizers need not be connected. Even though we are interested only in deformation rings with fixed multiplier, we have found it more convenient to carry out the analysis without fixing multipliers until the end. We also take advantage of the fact that we only care about GSp_4 (rather than, say, GSp_{2g}) to be a bit more ad hoc in our arguments.

Throughout the rest of this section, q will denote an integer which is not a multiple of p.

7.4.9. *Preliminaries on nilpotent matrices.* Let $\mathcal{U} \subset \mathrm{GSp}_4/\mathcal{O}$ be the closed subscheme of matrices with characteristic polynomial $(X-1)^4$, and let $\mathcal{N} \subset \operatorname{Lie}(\mathrm{GSp}_4)$ be the closed subscheme of matrices with characteristic polynomial X^4.

In [Tay08], under the assumption that $p \geq n$, Taylor uses truncations to degree X^{n-1} of the usual exponential and logarithm maps in order to relate unipotent and nilpotent matrices (see in particular [Tay08, Lem. 2.4]). For $p > 3$, we could in the same way use the truncations to order X^3 of the usual exponential and logarithmic maps. However, both exp and log to third order involve terms of the form $X^3/3!$ and $(X-1)^3/3$, which we need to avoid when working in residue characteristic three. In the proof of [Tho12, Lem. 3.15] an alternative approach is given (again in the case of GL_n), using the maps $\exp_1 = 1 + N$ and $\log_1 = (U-1)$ in order to avoid assumptions on the characteristic. However, neither the matrices $I + N$ for nilpotent N nor $U - 1$ for unipotent U will in general be symplectic, and thus our truncated exponential and logarithm maps must be at least quadratic. This motivates the following definitions.

For $p \geq 3$, we have the following modified versions of the exponential and logarithm map, which are the same as the usual definitions up to and including order X^2:

$$\begin{aligned} \exp_2 : \mathcal{N} &\to \mathcal{U} \\ N &\mapsto I + N + \frac{N^2}{2} + \frac{N^3}{2} \end{aligned}$$

and

$$\begin{aligned} \log_2 : \mathcal{U} &\to \mathcal{N} \\ U &\mapsto (U - I) - \frac{(U-I)^2}{2}. \end{aligned}$$

It is easily verified that these maps do indeed have image $\mathcal{U}$, respectively $\mathcal{N}$, and that they are in fact inverses to each other, and in particular are bijective. Additionally, they commute with the conjugation action of GSp_4, and for $m \in \mathbf{Z}$ satisfy

$$\exp_2\left(mN + m^* N^3\right) = \exp_2(N)^m, \quad \log_2(U^m) = m \log_2(U) + m^* \log_2(U)^3,$$

where $m^* = (m - m^3)/3 \in \mathbf{Z}$.

We define the following elements of $\mathcal{N}(\mathcal{O})$:

$$N_0 = 0,\ N_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ N_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ N_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We also let $\mathcal{N}_i \subset \mathcal{N}$ be the reduced, locally closed subscheme consisting of nilpotent matrices of rank i, so that $N_i \in \mathcal{N}_i$.

The following is an analogue of [Tay08, Lem. 2.5].

Proposition 7.4.10.

(1) $Z_{\mathrm{GSp}_4}(N_i)$ is a smooth group scheme over $\operatorname{Spec}\mathcal{O}$ with fibres of dimensions $11, 7, 5, 3$ for $i = 0, 1, 2, 3$. Each connected component of $Z_{\mathrm{GSp}_4}(N_i)$ is irreducible with irreducible special fibre. Moreover, $Z_{\mathrm{GSp}_4}(N_i)$ is irreducible except when $i = 2$, in which case it has two components.

(2) Locally in the étale topology, the universal nilpotent matrix over $\mathcal{N}_i$ is conjugate to N_i by a section of GSp_4.

(3) $\mathcal{N}_i$ is smooth over $\operatorname{Spec}\mathcal{O}$ with irreducible fibres of dimensions $0, 4, 6, 8$ for $i = 0, 1, 2, 3$. In particular, $\mathcal{N}_i$ is irreducible.

Proof. Part (1) can be checked by brute force calculation. For instance in the most interesting case when $i = 2$ a direct computation (using that $p > 2$) shows that

$$Z_{\mathrm{GSp}_4}(N_2) \simeq \mathcal{O}[x, y, z, w, \alpha, \beta, \gamma, \delta, (wx - yz)^{-1}]/(xy, wz, y\gamma - w\alpha - x\delta - z\beta)$$

where the matrix is given by

$$\begin{pmatrix} x & y & \alpha & \beta \\ z & w & \gamma & \delta \\ 0 & 0 & x & y \\ 0 & 0 & z & w \end{pmatrix}$$

and from this all the properties are clear (for instance, the two components are given by $x = w = 0$ and $y = z = 0$).

For part (2), we explain the case when $i = 2$. The others are similar but easier. We may view the universal nilpotent N over $\mathcal{N}_2$ as an endomorphism of $\mathcal{O}^4_{\mathcal{N}_2}$ with the "standard" symplectic form ψ. Then, by the definition of $\mathcal{N}_2$, $\ker(N)$ is a local direct summand of rank 2. Then one checks that

$$\begin{aligned} \psi' : \mathcal{O}^4_{\mathcal{N}_2}/\ker N \times \mathcal{O}^4_{\mathcal{N}_2}/\ker N &\to \mathcal{O}_{\mathcal{N}_2} \\ (v, w) &\mapsto \psi(Nv, w) \end{aligned}$$

is a well defined non-degenerate symmetric pairing.

Étale locally, one may trivialize ψ': For any point $x \in \mathcal{N}_2$ we may pick a Zariski open neighbourhood $x \in U = \operatorname{Spec} A \subset \mathcal{N}_2$ over which $\mathcal{O}^4_{\mathcal{N}_2}/\ker N$ has a basis f_1, f_2 with $\psi'(f_1, f_2) = 0$ and $\psi'(f_1, f_1), \psi'(f_2, f_2) \in A^\times$. Let A' be the étale A-algebra $A[\sqrt{\psi'(f_1, f_1)}, \sqrt{\psi'(f_2, f_2)}]$, so that over $U' = \operatorname{Spec} A'$, $(\mathcal{O}^4_{\mathcal{N}_2}/\ker N)_{U'}$ has a basis $f_1' = f_1/\sqrt{\psi'(f_1, f_1)}$, $f_2' = f_2/\sqrt{\psi'(f_2, f_2)}$ with $\psi'(f_1', f_1') = \psi'(f_2', f_2') = 1$ and $\psi'(f_1', f_2') = 0$. Now lift f_1' and f_2' to sections e_1 and e_2 of $\mathcal{O}^4_{U'}$. We may further arrange that $\psi(e_1, e_2) = 0$ by replacing e_2 by $e_2 - \psi(e_1, e_2)Ne_1$. Then Ne_2, Ne_1, e_1, e_2 forms a symplectic basis for $\mathcal{O}^4_{U'}$, and if we let $g \in \mathrm{GSp}_4$ have these elements as columns, then $N_{U'} = gN_2g^{-1}$.

Finally we turn to part (3). For each i, there is a map

$$\begin{aligned} \mathrm{GSp}_4 &\to \mathcal{N}_i \\ g &\mapsto gN_ig^{-1}. \end{aligned}$$

By the first two parts of the proposition, this map is smooth and surjective. Indeed, it suffices to check this after base change to a suitable étale cover $U \to \mathcal{N}_i$, over which it becomes isomorphic to $Z_{\mathrm{GSp}_4}(N_i)_U \to U$. It follows that $\mathcal{N}_i$ is smooth over $\mathcal{O}$. The fibres of $\mathcal{N}_i$ are irreducible because those of GSp_4 are, and the statement about dimensions follows from the computation of the dimensions of the fibres of $\mathrm{GSp}_4 \to \mathcal{N}_i$ in part (1). □

Remark 7.4.11. By contrast to the situation for GL_n considered in [Tay08], it is no longer the case that $Z_{\mathrm{GSp}_4}(N_i)$ is connected, nor is it true that the universal matrix over $\mathcal{N}_i$ is Zariski locally conjugate to N_i (both fail when $i = 2$).

7.4.12. *Some spaces of polynomials.* Let $\tilde{\mathcal{P}} = \mathbf{G}_m^3$ be the diagonal torus in GSp_4; we somewhat abusively write

$$\tilde{\mathcal{P}} = \{(X-\alpha)(X-\beta)(X-\gamma\beta^{-1})(X-\gamma\alpha^{-1})\}$$

where the order of the linear factors matters, and we let

$$\mathcal{P} = \tilde{\mathcal{P}}/W = \{X^4 + a_3X^3 + a_2X^2 + a_1X + a_0 \mid a_0 \in \mathbf{G}_m, a_3^2a_0 = a_1^2\},$$

so that there is a finite map $\pi : \tilde{\mathcal{P}} \to \mathcal{P}$, given by multiplying out the linear factors. We consider some reduced closed subspaces of $\mathcal{P}$:

$$\begin{aligned} \mathcal{P}_0 &= \mathcal{P} \\ \mathcal{P}_1 &= \pi(\{(X-\alpha)(X-\beta)(X-\gamma\beta^{-1})(X-\gamma\alpha^{-1}) \mid \gamma\alpha^{-1} = q\alpha\}) \\ \mathcal{P}_2 &= \pi(\{(X-\alpha)(X-q\alpha)(X-\gamma q^{-1}\alpha^{-1})(X-\gamma\alpha^{-1})\}) \\ \mathcal{P}_3 &= \pi(\{(X-\alpha)(X-q\alpha)(X-q^2\alpha)(X-q^3\alpha)\}) \end{aligned}$$

We will find it useful to consider some explicit elements of $\mathrm{GSp}_4(R)$, for an $\mathcal{O}$-algebra R. For $\alpha, \beta, \gamma \in R^\times$ we let

$$\begin{aligned} \Phi_0(\alpha,\beta,\gamma) &= \mathrm{diag}(\alpha,\beta,\gamma\beta^{-1},\gamma\alpha^{-1}) \\ \Phi_1(\alpha,\beta) &= \mathrm{diag}(q\alpha,\beta,q\alpha^2/\beta,\alpha) \\ \Phi_{2,a}(\alpha,\gamma) &= \mathrm{diag}(q\alpha,\gamma\alpha^{-1},\alpha,\gamma q^{-1}\alpha^{-1}) \\ \Phi_{2,b}(\alpha,\beta) &= \begin{pmatrix} 0 & q\alpha & 0 & 0 \\ q\beta & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & \beta & 0 \end{pmatrix} \\ \Phi_3(\alpha) &= \begin{pmatrix} q^3\alpha & 0 & \frac{q(1-q^2)}{6}\alpha & 0 \\ 0 & q^2\alpha & 0 & \frac{(1-q^2)}{6}\alpha \\ 0 & 0 & q\alpha & 0 \\ 0 & 0 & 0 & \alpha \end{pmatrix} \end{aligned}$$

7.4.13. *Spaces of matrices.* We define $\mathcal{N}(q)$ to be the closed subscheme of $\mathrm{GSp}_4 \times \mathcal{N}$ consisting of pairs (Φ, N) satisfying

$$\Phi N \Phi^{-1} = \log_2(\exp_2(N)^q) = qN + q^* N^3,$$

where as above we write $q^* = (q - q^3)/3$. This definition is motivated by the following. The actual equation we wish to study has the form

$$\Phi U \Phi^{-1} = U^q$$

for a unipotent matrix U. If we let $N = \log_2(U)$, we have $U = \exp_2(N)$, and so, applying $\log_2$ to the equation above, one finds precisely that

$$\Phi N \Phi^{-1} = \log_2(U^q) = \log_2(\exp_2(N)^q).$$

Noting that

$$N = \frac{1}{q}(qN + q^* N^3) - \frac{q^*}{q^4}(qN + q^* N^3)^3,$$

we see that the centralizers of N and $qN + q^* N^3$ coincide. It follows that if (Φ, N) is a point of $\mathcal{N}(q)$, then (Ψ, N) is another point if and only if $\Psi\Phi^{-1}$ centralizes N if and only if $\Psi^{-1}\Phi$ centralizes N. Note also that if $N^3 = 0$, then the condition on Φ is simply that $\Phi N \Phi^{-1} = qN$, while, if $q = 1$, then the equation is simply that $\Phi N \Phi^{-1} = N$.

Consider the projection

$$\begin{aligned} \mathcal{N}(q) &\to \mathcal{N} \\ (\Phi, N) &\mapsto N \end{aligned}$$

and let $\mathcal{N}(q)_i$ denote the locally closed preimage of $\mathcal{N}_i$. We let $\mathcal{Z}_i/\mathcal{N}_i$ be the centralizer of the universal element over $\mathcal{N}_i$. Then there is an action of $\mathcal{Z}_i$ on $\mathcal{N}(q)_i$ by $z \cdot (\Phi, N) = (z\Phi, N)$.

Proposition 7.4.14. *The above action makes $\mathcal{N}(q)_i$ into a $\mathcal{Z}_i$-torsor over $\mathcal{N}_i$.*

Proof. By Proposition 7.4.10, we may check the proposition after base change to a suitable étale cover $U \to \mathcal{N}_i$, over which the universal nilpotent over U is of the form gN_ig^{-1} for some $g \in \mathrm{GSp}_4(U)$. Let Φ_i be any of the explicit choices of Φ given above for N_i (for $i = 2$, take any specialization of either $\Phi_{2,a}$ or $\Phi_{2,b}$). Then one readily checks that (Φ_i, N_i) is a point on $\mathcal{N}(q)_i$, and that

$$\begin{aligned} (\mathcal{Z}_i)_U &\to (\mathcal{N}(q)_i)_U \\ z &\mapsto (zg\Phi_i g^{-1}, gN_ig^{-1}) \end{aligned}$$

is an isomorphism compatible with the $\mathcal{Z}_i$-action. □

Corollary 7.4.15. *For $i = 0, 1, 2, 3$, $\mathcal{N}(q)_i$ is smooth over $\mathcal{O}$ with fibres equidimensional of dimension 11. For $i \neq 2$, $\mathcal{N}(q)_i$ is irreducible with nonempty irreducible special fibre, while $\mathcal{N}(q)_2$ has two connected components, each of which is irreducible with nonempty irreducible special fibre.*

Proof. The smoothness and dimension are an immediate consequence of Propositions 7.4.10 and 7.4.14. Moreover, for $i \neq 2$, $\mathcal{N}(q)_i \to \mathcal{N}_i$ is flat with irreducible fibres, and $\mathcal{N}_i$ is irreducible, and hence $\mathcal{N}(q)_i$ is irreducible. The same argument applies to the special fibre.

Now we explain why $\mathcal{N}(q)_2$ has two connected components. As we explained in the proof of Proposition 7.4.10, over $\mathcal{N}_2$ we have the rank 2 non-degenerate

quadratic space $\mathcal{O}^4_{\mathcal{N}_2}/\ker(N)$ with quadratic form given by $v \mapsto \psi(v, Nv)$. Over $\mathcal{N}_2$ we have $N^3 = 0$, so the relation $\Phi N = qN\Phi$ holds on $\mathcal{N}(q)_2$, which implies that Φ preserves $\ker(N)$ and the computation

$$\psi(\Phi v, N\Phi v) = q^{-1}\psi(\Phi v, \Phi N v) = q^{-1}\nu(\Phi)\psi(v, Nv)$$

shows that Φ is an element of the general orthogonal group of this quadratic space.

This general orthogonal group has two components (corresponding to whether the determinant and multiplier agree or differ by a sign). As a result we may write $\mathcal{N}(q)_2 = \mathcal{N}(q)_{2,a} \coprod \mathcal{N}(q)_{2,b}$ where $\mathcal{N}(q)_{2,a}$ is the locus where Φ lies in the identity component and $\mathcal{N}(q)_{2,b}$ is the locus where Φ lies in the nonidentity component. Each of these loci is in fact nonempty; for example, we can consider points of the form $(\Phi_{2,a}(\alpha, \gamma), N_2)$ and $(\Phi_{2,b}(\alpha, \beta), N_2)$. As $\mathcal{N}(q)_{2,a}$ and $\mathcal{N}(q)_{2,b}$ are unions of connected components, the action of $\mathcal{Z}_2$ restricts to an action of the identity component $\mathcal{Z}_2^\circ$ on each of them, and one easily checks that they must each be torsors for $\mathcal{Z}_2^\circ$, and so the same argument as above shows that $\mathcal{N}(q)_{2,a}$ and $\mathcal{N}(q)_{2,b}$ are irreducible with nonempty irreducible special fibre. □

For the rest of this section, we will continue to use the notation $\mathcal{N}(q)_{2,a}$ and $\mathcal{N}(q)_{2,b}$ for the two connected components of $\mathcal{N}(q)_2$ as introduced in the proof of Corollary 7.4.15. We also write $\overline{\mathcal{N}(q)}_i$ for the Zariski closure of $\mathcal{N}(q)_i$, $\overline{(\mathcal{N}(q)_{i,\mathbf{F}})}$ for the Zariski closure of its special fibre, and so on.

Proposition 7.4.16. *The irreducible components of $\mathcal{N}(q)$ are $\overline{\mathcal{N}(q)}_{2,a}$, $\overline{\mathcal{N}(q)}_{2,b}$, and $\overline{\mathcal{N}(q)}_i$ for $i = 0, 1, 3$. The irreducible components of the special fibre $\mathcal{N}(q)_{\mathbf{F}}$ are $\overline{(\mathcal{N}(q)_{2,a,\mathbf{F}})}$, $\overline{(\mathcal{N}(q)_{2,b,\mathbf{F}})}$, and $\overline{(\mathcal{N}(q)_{i,\mathbf{F}})}$ for $i = 0, 1, 3$. Each irreducible component of $\mathcal{N}(q)$ has irreducible and generically reduced special fibre.*

Proof. $\mathcal{N}(q)$ is set theoretically the disjoint union of the five locally closed subschemes $\mathcal{N}(q)_{2,a}$, $\mathcal{N}(q)_{2,b}$, and $\mathcal{N}(q)_i$ for $i = 0, 1, 3$, which are each irreducible and of the same dimension by Corollary 7.4.15. Hence their closures are the irreducible components of $\mathcal{N}(q)$. The same argument applies to the special fibre.

To prove the last statement it will suffice to prove that for $i = 0, 1, 2, 3$, $\overline{\mathcal{N}(q)}_i$ does not contain the generic points of $\mathcal{N}(q)_{j,\mathbf{F}}$ for $j \neq i$. Indeed it already follows from Corollary 7.4.15 that $(\overline{\mathcal{N}(q)_{2,a}})_{\mathbf{F}}$ does not contain the generic point of $\mathcal{N}(q)_{2,b,\mathbf{F}}$ and vice versa; and we also see that the special fibre of each irreducible component of $\mathcal{N}(q)$ is reduced at the generic point of the corresponding component of $\mathcal{N}(q)_{\mathbf{F}}$.

In order to do this for $i = 0, 1, 2, 3$, let $\widetilde{\mathcal{N}(q)}_i \subset \mathcal{N}(q)$ be the reduced closed subscheme consisting of pairs (Φ, N) such that $\mathrm{rank}(N) \leq i$ and the characteristic polynomial $\mathrm{char}_\Phi(X)$ is in $\mathcal{P}_i$. An easy calculation shows that $\mathcal{N}(q)_i \subset \widetilde{\mathcal{N}(q)}_i$, and hence $\overline{\mathcal{N}(q)}_i \subset \widetilde{\mathcal{N}(q)}_i$. (One can either follow the proof of [Tho12, Lem. 3.15], or observe that we have seen above that it is enough to check that this holds for the points of the form $(z_i\Phi_i, N_i)$ for our explicit choices of Φ_i and for $z_i \in \mathcal{Z}_i$.) Thus to conclude the proof, all we have to do is exhibit a point on each irreducible component of $\mathcal{N}(q)_{\mathbf{F}}$ which is only contained in one of the $\widetilde{\mathcal{N}(q)}_i$'s. For instance, we may take the following five points:

- $(\Phi_0(\alpha, \beta, \gamma), 0)$ for general values of $\alpha, \beta, \gamma \in \overline{\mathbf{F}}^\times$.
- $(\Phi_1(\alpha, \beta), N_1)$ for general values of $\alpha, \beta \in \overline{\mathbf{F}}^\times$.

- $(\Phi_{2,a}(\alpha,\gamma), N_2)$ for general values of $\alpha,\gamma \in \overline{\mathbf{F}}^\times$.
- $(\Phi_{2,b}(\alpha,\beta), N_2)$ for general values of $\alpha,\beta \in \overline{\mathbf{F}}^\times$.
- $(\Phi_3(1), N_3)$. □

For $x, y \in \mathcal{O}^\times$ and q a positive integer which is not a multiple of p, we let $\mathcal{M}(x,y;q)$ be the closed subscheme of $\mathrm{GSp}_4^2/\mathcal{O}$ consisting of pairs (Φ,Σ) satisfying:

- The characteristic polynomial of Σ is $(X-x)(X-y)(X-y^{-1})(X-x^{-1})$.
- $\Phi\Sigma\Phi^{-1} = \Sigma^q$.

We note that the order of x and y doesn't matter.

There is evidently an isomorphism

$$\begin{aligned}\mathcal{M}(1,1;q) &\to \mathcal{N}(q)\\ (\Phi,\Sigma) &\mapsto (\Phi, \log_2(\Sigma)).\end{aligned}$$

We now have the following analogue of [Tay08, Lem. 3.2].

Proposition 7.4.17. *Let q be a positive integer with $q \equiv 1 \pmod p$.*

(1) Let $\mathcal{M}_i$ be the irreducible components of $\mathcal{M}(1,1;q)$ with their reduced subscheme structure. Then the special fibres $\mathcal{M}_{i,\mathbf{F}}$ are distinct, generically reduced and irreducible, and their reductions are precisely the irreducible components of $\mathcal{M}(1,1;q)_{\mathbf{F}}$.

(2) Suppose that either $q \neq 1$ and x, y are non trivial $(q-1)$st roots of 1 in $1+\lambda\mathcal{O}$ with $x \neq y^{\pm 1}$; or that $q = 1$ and x, y are arbitrary elements of $1+\lambda\mathcal{O}$. Then $\mathcal{M}(x,y;q)^{\mathrm{red}}$ is flat over $\mathcal{O}$.

Proof.

(1) This is an immediate consequence of Proposition 7.4.16 and the isomorphism $\mathcal{M}(1,1;q) \simeq \mathcal{N}(q)$ above.

(2) When $q \neq 1$, we observe that, as x, y, y^{-1}, x^{-1} are distinct $(q-1)$st roots of unity,

$$\mathrm{char}_\Sigma(X) = (X-x)(X-y)(X-x^{-1})(X-y^{-1}) | (X^{q-1}-1).$$

Hence, by the Cayley–Hamilton theorem, $\Sigma^q = \Sigma$. This implies that there is an isomorphism $\mathcal{M}(x,y;q) = \mathcal{M}(x,y;1)$. We are therefore reduced to the case that $q = 1$.

To show that $\mathcal{M}(x,y;1)^{\mathrm{red}}$ is flat over $\mathcal{O}$, it suffices to show that each generic point of its special fibre is the specialization of a point of the generic fibre. It suffices in turn to show that a Zariski dense set of points of the special fibre lift to the generic fibre. Then as x and y reduce to 1, we have $\mathcal{M}(x,y;1)_{\mathbf{F}} = \mathcal{M}(1,1;1)_{\mathbf{F}} \simeq \mathcal{N}(1)_{\mathbf{F}}$. This isomorphism, combined with the proof of Proposition 7.4.16, shows that the following five kinds of $\overline{\mathbf{F}}$-points are Zariski dense in $\mathcal{M}(x,y;1)_{\mathbf{F}}$ (because the corresponding points are dense in each $\widetilde{\mathcal{N}(q)}_i$):

- $(g\Phi_0(\alpha,\beta,\gamma)g^{-1}, 1)$
- $(g\Phi_1(\alpha,\beta)g^{-1}, g\exp_2(N_1)g^{-1})$
- $(g\Phi_{2,a}(\alpha,\gamma)g^{-1}, g\exp_2(N_2)g^{-1})$
- $(g\Phi_{2,b}(\alpha,\beta)g^{-1}, g\exp_2(N_2)g^{-1})$
- $(g\Phi_3(\alpha)g^{-1}, g\exp_2(N_3)g^{-1})$

where $\alpha, \beta, \gamma \in \overline{\mathbf{F}}^\times$ and $g \in \mathrm{GSp}_4(\overline{\mathbf{F}})$. Then letting $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} \in W(\overline{\mathbf{F}})$ and $\tilde{g} \in \mathrm{GSp}_4(W(\overline{\mathbf{F}}))$ be lifts, we can lift these to $W(\overline{\mathbf{F}})$ points of $\mathcal{M}(x, y; 1)$ of the following form (recall that we are in the case $q = 1$):

- $(\tilde{g}\Phi_0(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})\tilde{g}^{-1}, \tilde{g}\mathrm{diag}(x, y, y^{-1}, x^{-1})\tilde{g}^{-1})$
- $(\tilde{g}\Phi_1(\tilde{\alpha}, \tilde{\beta})\tilde{g}^{-1}, \tilde{g}\mathrm{diag}(x, y, y^{-1}, x^{-1}) \exp_2(N_1)\tilde{g}^{-1})$
- $(\tilde{g}\Phi_{2,a}(\tilde{\alpha}, \tilde{\gamma})\tilde{g}^{-1}, \tilde{g}\mathrm{diag}(x, y, y^{-1}, x^{-1}) \exp_2(N_2)\tilde{g}^{-1})$
- $(\tilde{g}\Phi_{2,b}(\tilde{\alpha}, \tilde{\beta})\tilde{g}^{-1}, \tilde{g}\mathrm{diag}(A, X(A^\tau)^{-1}X) \exp_2(N_2)\tilde{g}^{-1})$, where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and A is a 2 by 2 matrix with coefficients in $W(\overline{\mathbf{F}})$ which has trivial reduction, commutes with $\begin{pmatrix} 0 & \tilde{\alpha} \\ \tilde{\beta} & 0 \end{pmatrix}$ and has eigenvalues x, y (for the existence of such a matrix, use that $\begin{pmatrix} 0 & \tilde{\alpha} \\ \tilde{\beta} & 0 \end{pmatrix}$ has distinct eigenvalues mod p, and is therefore diagonalizable).
- $(\tilde{g}\Phi_3(\tilde{\alpha})\tilde{g}^{-1}, \tilde{g}\mathrm{diag}(x, y, y^{-1}, x^{-1}) \exp_2(N_3)\tilde{g}^{-1})$. □

Next we have an analogue of [Tay08, Lem. 3.4].

Proposition 7.4.18. *Let $q > 1$ with $q \equiv 1 \pmod p$ and let x, y be non trivial $(q-1)$st roots of 1 in $1 + \lambda\mathcal{O}$ with $x \neq y^{\pm 1}$. Let $R = \hat{\mathcal{O}}_{\mathcal{M}(x,y;q),(1,1)}$ be the complete local ring of $\mathcal{M}(x, y; q)$ at the point $(1, 1)$ of the special fibre. Then $\mathrm{Spec}\, R[1/p]$ is connected.*

Proof. The proof of [Tay08, Lem. 3.4] carries over with minor modifications. Let $\wp_0$ denote the maximal ideal of $R[1/p]$ corresponding to (Φ_0, Σ_0) with Φ_0 trivial and Σ_0 the diagonal matrix $\mathrm{diag}(x, y, x^{-1}, y^{-1})$, and let $\wp$ be another maximal ideal, corresponding to a pair (Φ, Σ). We need to show that $\wp$ is in the same connected component as $\wp_0$. One deduces as in [Tay08] that $\wp$ is in the same connected component of $\mathrm{Spec}(R[1/p])$ as the maximal ideal corresponding to $(E^{-1}\Phi E, E^{-1}\Sigma E)$ where $E \in \mathrm{GSp}_4(\mathcal{O})$ is arbitrary. In order to pass to an upper triangular form, we require the existence of a filtration Fil^i of $k(\wp)^4$ such that:

(1) Each Fil^i is preserved by Φ and Σ.
(2) The graded pieces gr^i are one dimensional and their eigenvalues (in order) are α, β, $\gamma\beta^{-1}$, $\gamma\alpha^{-1}$, which are the generalized eigenvalues of Φ.
(3) The orthogonal complement of Fil^i is Fil^{4-i}.

As in the proof of Proposition 7.4.17, Φ and Σ commute, so we may choose Fil^1 to be a common eigenvector of Φ and Σ. We define Fil^3 to be the orthogonal complement of Fil^1, and then choose Fil^2 to be any lift of a common eigenvector of Φ and Σ in $\mathrm{Fil}^3 / \mathrm{Fil}^1$.

The constructions of paths in [Tay08] from upper triangular to diagonal and between diagonal matrices (eventually to (Φ_0, Σ_0) and thus connecting $\wp$ to $\wp_0$) have obvious symplectic modifications. □

7.4.19. *Application to deformation rings.* Now let $\chi = (\chi_1, \chi_2)$ be a pair of continuous characters $\chi_i : \mathcal{O}_{F_v}^\times \to \mathcal{O}^\times$ that are trivial mod λ, let $\tilde{\mathcal{D}}_v^\chi$ be the functor on $\mathrm{CNL}_\mathcal{O}$ of continuous homomorphisms $\rho : G_{F_v} \to \mathrm{GSp}_4(A)$ which are trivial mod $\mathfrak{m}_A$ and such that for $\sigma \in I_{F_v}$, the characteristic polynomial of $\rho(\sigma)$is

$$(X - \chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1})(X - \chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}).$$

As in §7.4, we let $\mathcal{D}_v^\chi \subset \tilde{\mathcal{D}}_v^\chi$ be the subfunctor of ρ with $\nu \circ \rho = \varepsilon^{-1}$. The functors $\tilde{\mathcal{D}}_v^\chi$ and $\mathcal{D}_v^\chi$ are representable by rings $\tilde{R}_v^\chi$ and R_v^χ. We also let $\mathcal{D}_1$ be the functor with $\mathcal{D}_1(A)$ parameterizing continuous unramified characters $\psi : G_{F_v} \to A^\times$ which are trivial mod $\mathfrak{m}_A$. It is representable by $\mathcal{O}[[T]]$ with universal object $\chi^{\mathrm{univ}} : G_{F_v} \to \mathcal{O}[[T]]^\times$ given by $\chi^{\mathrm{univ}}(\mathrm{Frob}_v) = 1 + T$.

For any $A \in \mathrm{CNL}_\mathcal{O}$, then as A is complete and $p > 2$,

$$\begin{aligned} 1 + \mathfrak{m}_A &\to 1 + \mathfrak{m}_A \\ t &\mapsto t^2 \end{aligned}$$

is a bijection and we denote its inverse by $x \mapsto \sqrt{x}$. Then we have

Proposition 7.4.20. *There is an isomorphism of functors*

$$\begin{aligned} \mathcal{D}_v^\chi \times \mathcal{D}_1 &\to \tilde{\mathcal{D}}_v^\chi \\ (\rho, \psi) &\mapsto \rho \otimes \psi \end{aligned}$$

Consequently there is an isomorphism $R_v^\chi[[T]] \simeq \tilde{R}_v^\chi$.

Proof. For the inverse we may take the natural transformation

$$\begin{aligned} \tilde{\mathcal{D}}_v^\chi &\to \mathcal{D}_v^\chi \times \mathcal{D}_1 \\ \rho &\mapsto (\rho \otimes \sqrt{\varepsilon \cdot (\nu \circ \rho)}^{-1}, \sqrt{\varepsilon \cdot (\nu \circ \rho)}) \end{aligned}$$

The only thing that we need to check is that if $\rho \in \tilde{\mathcal{D}}_v^\chi(A)$ then $\nu \circ \rho$ is trivial on I_{F_v}. For $\sigma \in I_{F_v}$, $(\nu \circ \rho(\sigma))^2$ is the constant term of the characteristic polynomial of $\rho(\sigma)$ which is 1 by definition. But also $\nu \circ \rho(\sigma) \equiv 1 \pmod{\mathfrak{m}_A}$, and hence $\nu \circ \rho(\sigma) = 1$ as $p > 2$. □

We may now relate these deformation rings to the spaces of matrices considered in this section.

Proposition 7.4.21. *Let σ be a chosen topological generator of the tame inertia subgroup of G_{F_v}. Let $x = \chi_1(\mathrm{Art}_{F_v}^{-1}(\sigma))$ and $y = \chi_2(\mathrm{Art}_{F_v}^{-1}(\sigma))$. Then*

$$\tilde{R}_v^\chi \simeq \hat{\mathcal{O}}_{\mathcal{M}(x,y;q_v),(1,1)}.$$

Proof. Since $\overline{\rho}|_{G_{F_v}}$ is trivial, any lifting of it factors through the quotient $T_v = G_{F_v}/P_{F_v}$, where P_{F_v} denotes the maximal pro-prime-to-p subgroup of I_{F_v} (that is, the kernel of any non-trivial homomorphism $I_{F_v} \to \mathbf{Z}_p$). If φ is an arithmetic Frobenius element in G_{F_v}, then the group T_v is topologically generated by φ and the image of σ, subject to the constraints that σ generates a pro-p group, and that $\varphi\sigma\varphi^{-1} = \sigma^{q_v}$. The result then follows from the definitions. □

We can now conclude the proofs of Propositions 7.4.7 and 7.4.8 exactly as in [Tay08].

Proof of Proposition 7.4.7. Combining Propositions 7.4.17 (1) and 7.4.21 with [Tay08, Lem. 2.7] proves the corresponding result for $\tilde{R}_v^1$. The result for R_v^1 follows from this and Proposition 7.4.20. □

Proof of Proposition 7.4.8. Proposition 7.4.17 implies that $(\tilde{R}_v^\chi)^{\mathrm{red}}$ is flat over $\mathcal{O}$. Proposition 7.4.18 implies that $\mathrm{Spec}(\tilde{R}_v^\chi[1/p])$ is connected. On the other hand, by Lemma 7.4.6, for any closed point $x \in \mathrm{Spec}(\tilde{R}_v^\chi[1/p])$, the localization $(\tilde{R}_v^\chi[1/p])_x$

is regular and hence a domain. Then the result follows from Propositions 7.4.17 (2) and 7.4.21, as in the proof of [Tay08, Prop. 3.1]. □

7.5. **Big image conditions and vast representations.**

7.5.1. *Enormous subgroups.* Following [CG18, KT17] (which give the analogous definition for GL_n) we now define the notion of "enormous image," with some minor modifications.

Definition 7.5.2. We say that a subgroup $H \subset \mathrm{GSp}_4(k)$ is *enormous* if it satisfies the following conditions:

(E1) $H^1(H, \mathrm{ad}^0) = 0$ for the 10-dimensional representation ad^0.
(E2) H acts absolutely irreducibly in its natural representation, in particular, $H^0(H, \mathrm{ad}^0) = 0$.
(E3) For all simple $\overline{k}[H]$-submodules $W \subset \overline{k} \otimes \mathrm{ad}^0$, there is an element $h \in H$ such that
  - $h \in \mathrm{GSp}_4(k)$ has 4 distinct eigenvalues, and
  - 1 is an eigenvalue for the action of h on W.

If H only satisfies (E2) and (E3), then we say that H is *weakly enormous.*

Lemma 7.5.3. *If H and H' are subgroups of $\mathrm{GSp}_4(k)$ with the same image in $\mathrm{PGSp}_4(k)$, then H is enormous (resp. weakly enormous) if and only if H' is enormous (resp. weakly enormous).*

Proof. Suppose that P is the projective image of H in $\mathrm{PGSp}_4(k)$ and Z is the kernel. Then the action of H on ad^0 factors through P. In particular, $H^0(P, \mathrm{ad}^0) = H^0(H, \mathrm{ad}^0)$, and there is an inflation–restriction sequence

$$0 \to H^1(P, \mathrm{ad}^0) \to H^1(H, \mathrm{ad}^0) \to H^1(Z, \mathrm{ad}^0)^P = 0.$$

Hence all the conditions in the definitions of enormousness and weakly enormousness depend only on the projective representation. □

Note that if $H' \subset H$ is weakly enormous, then so is H, but if H' is enormous, then H is not necessarily enormous.

Remark 7.5.4. Some "big image" conditions in the literature have the additional assumption that H has no p-power quotient. In practice, however, that hypothesis is often only used in a very weak way, namely, to ensure that the image of $\overline{\rho}$ restricted to $G_{F(\zeta_p)}$ coincides with the restriction to $G_{F(\zeta_{p^N})}$ for all $N \geq 1$. The stronger hypothesis has the unfortunate side effect of ruling out some perfectly fine Galois representations to which the Taylor–Wiles method applies, most notably, surjective representations $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_3)$ with cyclotomic determinant (exactly the case which arises in the original work of Wiles!). In order not to rule out some interesting subgroups which occur for $p = 3$, we therefore do not assume this hypothesis.

Let $p \geq 3$. The cyclotomic character induces a homomorphism:

$$G_F \to \mathbf{Z}_p^\times \simeq (\mathbf{Z}/p\mathbf{Z})^\times \oplus (1 + p\mathbf{Z}_p) \to (1 + p\mathbf{Z}_p).$$

If p is unramified in F, then this composite map is surjective. In general, the image contains $1 + p^\delta$ for some integer δ. In order to address the passage from $F(\zeta_p)$ to $F(\zeta_{p^N})$ in the Taylor–Wiles argument, we have the following lemma:

Lemma 7.5.5. *Suppose that $p \geq 3$. Let*

$$\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$$

be a continuous homomorphism. Then there exists an integer δ depending only on F such that the image of $\overline{\rho}$ restricted to $G_{F(\zeta_{p^N})}$ is independent of N for $N \geq 1 + \delta$ if $p \geq 5$ or $N \geq 2 + \delta$ for $p = 3$. If p is unramified in F, then one may take $\delta = 1$.

Proof. There is a canonical injective homomorphism

$$\mathrm{Gal}(F(\zeta_{p^N})/F) \to (\mathbf{Z}/p^N\mathbf{Z})^\times$$

for all N, and we will identify $\mathrm{Gal}(F(\zeta_{p^N})/F)$ with its image in $(\mathbf{Z}/p^N\mathbf{Z})^\times$ in the below. We choose δ such that for all N, the image contains $1 + p^\delta$. In particular, if p is unramified in F, we can take $\delta = 1$.

Let M denote the fixed field of $\overline{\rho}$. There are natural maps as follows:

$$\begin{tikzcd}
\mathrm{Gal}(M(\zeta_{p^{N+1}})/F) \arrow[r, hook] \arrow[d] & \mathrm{Gal}(M/F) \times \mathrm{Gal}(F(\zeta_{p^{N+1}})/F) \arrow[d] \\
\mathrm{Gal}(M(\zeta_{p^N})/F) \arrow[r, hook] & \mathrm{Gal}(M/F) \times \mathrm{Gal}(F(\zeta_{p^N})/F)
\end{tikzcd}$$

where the composites of the horizontal maps with the projections to each factor are surjective. The images of $\overline{\rho}$ restricted to $F(\zeta_{p^N})$ and $F(\zeta_{p^{N+1}})$ coincide precisely when the left hand vertical map has non-trivial kernel (necessarily of order p). We prove this is so under our assumptions on N.

It suffices to show that the horizontal image of the upper map contains an element of the form $(\mathrm{id}_M, 1 + mp^N)$ for some m with $(m, p) = 1$. By the surjectivity onto the second factor, it contains an element of the form $(g, 1 + p^\delta)$. Let m be the prime to p order of g, so that $h := g^m$ has p-power order. Since $p > 2$, we have $(g, 1 + p^\delta)^{mp^{N-\delta}} = (h^{p^{N-\delta}}, 1 + mp^N)$, and hence we are done providing the order of h divides $p^{N-\delta}$. Yet all p-power elements of $\mathrm{GSp}_4(k)$ have order dividing p if $p \geq 5$ or order dividing p^2 if $p = 3$. (The p-Sylow subgroup of $\mathrm{GSp}_4(k)$ consists of unipotent matrices which satisfy $(\sigma - 1)^4 = 0$, so $\sigma^{p^k} = 1$ when $p^k \geq 4$.) $\square$

In anticipation of Lemma 7.5.9 below, we make the following definition:

Definition 7.5.6. A representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$ is *vast* if one of the following two conditions holds:

(1) The image of $\overline{\rho}$ restricted to $G_{F(\zeta_{p^N})}$ is enormous for all sufficiently large N.
(2) The image of $\overline{\rho}$ restricted to $G_{F(\zeta_{p^N})}$ is weakly enormous for all sufficiently large N, and the fixed field L of $\mathrm{ad}^0\,\overline{\rho}$ does not contain ζ_p.

Remark 7.5.7. If p is unramified in F, then, in Definition 7.5.6, one may replace *sufficiently large* N by $N = 3$, since, by Lemma 7.5.5, the image in this case does not depend on N for $N \geq 3$.

Remark 7.5.8. By Lemma 7.5.3, $\overline{\rho}$ is vast if and only if any twist of $\overline{\rho}$ by a character is vast.

The following lemma will prove useful for constructing Taylor–Wiles primes:

Lemma 7.5.9. *Suppose that $p \geq 3$. Let $\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$ be a continuous representation. Fix an integer $N \geq 1$. Suppose either that:*

(1) The fixed field L of $\mathrm{ad}^0 \overline{\rho}$ does not contain ζ_p, or
(2) The restriction of $\overline{\rho}$ to $G_{F(\zeta_{p^N})}$ has enormous image.

Then

$$H^1(L(\zeta_{p^N})/F, \mathrm{ad}^0 \overline{\rho}(1)) = 0.$$

In particular, if $\overline{\rho}$ is vast, then the conclusion above holds for all sufficiently large N.

Proof. We first consider the case when $\zeta_p \notin L$. By inflation–restriction, it suffices to prove that the groups

$$H^1(L(\zeta_p)/F, \mathrm{ad}^0 \overline{\rho}(1)), \quad H^1(L(\zeta_{p^N})/L(\zeta_p), \mathrm{ad}^0 \overline{\rho}(1))^{\mathrm{Gal}(L(\zeta_p)/F)}$$

both vanish. The group $\mathrm{Gal}(L(\zeta_p)/L) \subset \mathrm{Gal}(L(\zeta_p)/F)$ acts trivially (by conjugation) on both the group $\mathrm{Gal}(L(\zeta_{p^N})/L(\zeta_p))$ and the module ad^0. However, it acts by non-trivial scalars on the twist $\mathrm{ad}^0(1)$ since we are assuming $\zeta_p \notin L$. Hence the second group vanishes after taking invariants. Applying inflation–restriction now to the first group, it suffices to prove that the groups

$$H^1(L/F, (\mathrm{ad}^0 \overline{\rho}(1))^{\mathrm{Gal}(L(\zeta_p)/L)}), \quad H^1(L(\zeta_p)/L, \mathrm{ad}^0 \overline{\rho}(1))^{\mathrm{Gal}(L/F)}$$

both vanish. The second group vanishes because $p \nmid [L(\zeta_p) : L]$. The first group vanishes because $\mathrm{ad}^0 \overline{\rho}$ is fixed by $\mathrm{Gal}(L(\zeta_p)/L)$ and thus has no invariants after being twisted by the mod-p cyclotomic character (which by assumption is a non-trivial character of $\mathrm{Gal}(L(\zeta_p)/L)$).

Now we consider the second case. Let M denote the splitting field of $\overline{\rho}$, so that M/L is a (possibly trivial) cyclic extension of degree prime to p. Inflation–restriction shows that we have an injection

$$H^1(L(\zeta_{p^N})/F, \mathrm{ad}^0 \overline{\rho}(1)) \hookrightarrow H^1(M(\zeta_{p^N})/F, \mathrm{ad}^0 \overline{\rho}(1)),$$

so it suffices to show that the latter group vanishes. By inflation–restriction, it is enough to show that the cohomology groups

$$H^1(F(\zeta_{p^N})/F, H^0(M(\zeta_{p^N}/F(\zeta_{p^N}), \mathrm{ad}^0 \overline{\rho}(1))),$$
$$H^1(M(\zeta_{p^N})/F(\zeta_{p^N}), \mathrm{ad}^0 \overline{\rho}(1))$$

both vanish. We are assuming that

$$H = \mathrm{Gal}(M(\zeta_{p^N})/F(\zeta_{p^N}))$$

is enormous. Thus to show that both groups above vanish, it suffices to note that

$$H^0(H, \mathrm{ad}^0) = H^1(H, \mathrm{ad}^0) = 0$$

because H is enormous. □

Remark 7.5.10. The two parts of this proof are essentially standard — in particular the first part is exactly the same as the proof of Lemma 5.3 of [Pil11].

We will require a weakly enormous (or in practice vast) image assumption in order to use the Cebotarev density theorem to guarantee the existence of Taylor–Wiles primes. Similarly, the following condition will allow us to use Cebotarev to arrange for our level structures to be neat by increasing the level at an auxiliary prime.

Definition 7.5.11. We say that a subgroup $H \subset \mathrm{GSp}_4(k)$ is *tidy* if there is an $h \in H$ with $\nu(h) \neq 1$, and such that no two eigenvalues of h have ratio $\nu(h)$ (but the eigenvalues need not be distinct). We say that a representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$ is tidy if it has tidy image.

Note that the property of tidiness is inherited from subgroups.

Lemma 7.5.12. *Suppose that $H \subset \mathrm{GSp}_4(k)$ is absolutely irreducible, and the centre Z of H has order at least 3. Then H is tidy.*

Proof. By Schur's lemma, the centre is cyclic and any element in the centre is scalar with eigenvalues $(\zeta, \zeta, \zeta, \zeta)$ for some ζ. If $|Z| \geq 3$, there thus exists such an element h in the centre with $\zeta^2 \neq 1$. Since $\nu(h) = \zeta^2 \neq 1$, and since the ratio of every pair of eigenvalues is $1 \neq \nu(h)$, it follows that H is tidy. □

Lemma 7.5.13. *Let $\Delta \subset \mathrm{GL}_2(\mathbf{F}_p) \times \mathrm{GL}_2(\mathbf{F}_p)$ be the subgroup of pairs (A, B) with $\det(A) = \det(B)$, and consider Δ as a subgroup of $\mathrm{GSp}_4(\mathbf{F}_p)$ via the map of §2.2. If $p \geq 5$ and $\Delta \subset H$, then H is tidy.*

Proof. The argument is very similar to the proof Lemma 7.5.12. The group Δ contains a cyclic subgroup of scalar matrices of order $p - 1 > 2$. □

Lemma 7.5.14. *If $p \geq 11$ and $H \subset \mathrm{GSp}_4(k)$ is absolutely irreducible, then conditions* (E1) *and* (E2) *are satisfied.*

Proof. This is immediate from [Tho12, Thm. A.9]. □

Lemma 7.5.15. *If $p \geq 3$, then $H = \mathrm{Sp}_4(\mathbf{F}_p)$ is enormous and $G = \mathrm{GSp}_4(\mathbf{F}_p)$ is tidy. If $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_p)$ is a surjective representation with similitude character $\overline{\varepsilon}^{-1}$, then $\overline{\rho}$ is vast and tidy.*

Proof. For all such p, the representation ad^0 is absolutely irreducible. Hence for weak enormity it suffices to note that H contains elements with distinct eigenvalues, and every such element has at least one eigenvalue 1 on ad^0. Thus for enormity it suffices to check that $H^1(\mathrm{Sp}_4(\mathbf{F}_p), \mathrm{ad}^0) = 0$. For $p \geq 11$, this follows from Lemma 7.5.14. For $p = 3$, 5, and 7, it can be checked directly using `magma` [BCP97]. (All of the `magma` code and output for this paper can be found at the github respository here [BCGP21].)

For tidiness, the centre of G has order $p-1$ so the result follows from Lemma 7.5.12 when $p > 3$. (It also follows from Lemma 7.5.13.) When $p = 3$, the group $\mathrm{GSp}_4(\mathbf{F}_3)$ contains an element g of order 20 with $\nu(g) = -1$; more precisely, its eigenvalues are of the form $\zeta, \zeta^3, \zeta^9, \zeta^{27}$ for a 20th root of unity ζ, and $\nu(g) = \zeta^{10} = \zeta^{30} = -1$. The ratios of the pairs of eigenvalues are of the form ζ^{3-1}, ζ^{9-1}, and ζ^{27-1}, and since none of these quantities is equal to $\zeta^{10} = -1$, we are done.

For vastness, note that the image of $\overline{\rho}$ restricted to $G_{F(\zeta_p)}$ will be $H = \mathrm{Sp}_4(\mathbf{F}_p)$. Since this group has no quotients of p-power order (indeed $\mathrm{PSp}_4(\mathbf{F}_p)$ is simple), the image of the restriction of $\overline{\rho}$ to $G_{F(\zeta_{p^N})}$ will also be H for all N. Hence the image of $\overline{\rho}$ restricted to $G_{F(\zeta_{p^N})}$ is always H and hence enormous; thus $\overline{\rho}$ is vast. □

7.5.16. *Representations Induced From Index Two Subgroups.* Suppose that $G \subset \mathrm{Sp}_4(k)$ is an absolutely irreducible subgroup such that the underlying representation W becomes reducible on an index two subgroup H. Write χ for the quadratic character $\chi : G \to G/H \to k^\times$ (we assume the characteristic of k is different

from 2). Write $G/H = \{1, \sigma\}$. Then one may write $W|_H = V \oplus V^\sigma$, and one has the following G-equivariant decompositions (not necessarily into irreducibles):

$$W \otimes W = W \otimes W^\vee = k \oplus k(\chi) \oplus \mathrm{Ind}_H^G(\mathrm{ad}^0(V)) \oplus \mathrm{As}(V) \oplus \mathrm{As}(V) \otimes \chi,$$
$$\mathrm{ad}^0(W) = \mathrm{Sym}^2(W) = \mathrm{Ind}_H^G(\mathrm{ad}^0(V)) \oplus \mathrm{As}(V),$$
$$\wedge^2(W) = k \oplus k(\chi) \oplus \mathrm{As}(V) \otimes \chi.$$

Here $\mathrm{As}(V)$ is the Asai representation, which satisfies $\mathrm{As}(V)|_H = V \otimes V^\sigma$. These identifications follow from computing what happens over H and noting that $W \simeq W \otimes \chi$.

Lemma 7.5.17. *Suppose that* $\mathrm{As}(V)$ *and* $\mathrm{Ind}_H^G(\mathrm{ad}^0(V))$ *are absolutely irreducible representations of* G*. Suppose that* $G\backslash H$ *has an element* g *of order neither dividing* 4 *nor divisible by* p*. Then* G *satisfies condition* (E3) *of enormousness.*

Proof. Let g be an element of $G\backslash H$. Since W is induced, the eigenvalues of g are invariant under multiplication by -1. Since $G \subset \mathrm{Sp}_4(k)$, the eigenvalues are invariant under inversion. It follows that the eigenvalues are of the form $(\alpha, \alpha^{-1}, -\alpha, -\alpha^{-1})$ for some α. If g has order neither dividing 4 nor divisible by p, then $\alpha^4 \neq 1$ and these eigenvalues are all distinct. To show (E3), it is enough to show that any such element g has an eigenvalue 1 on both $\mathrm{As}(V)$ and $\mathrm{Ind}_H^G(\mathrm{ad}^0(V))$. Let $\Gamma = \langle g \rangle$, and work in the Grothendieck group of representations of Γ. The representation Sym^2 differs from $\wedge^2$ by containing the squares of all the eigenvalues. Hence

$$[\mathrm{Sym}^2] = [\wedge^2] + [\alpha^2, \alpha^{-2}, \alpha^2, \alpha^{-2}].$$

Moreover, since $\chi(g) = -1$,

$$[\wedge^2] = [1] + [-1] + [-\mathrm{As}(V)].$$

It follows by counting eigenvalues in $W \otimes W$ that

$$[\wedge^2] = [1] + [-1] + [1, -1, -\alpha^2, -\alpha^{-2}],$$
$$[\mathrm{Sym}^2] = [-1, 1, \alpha^2, \alpha^{-2}] + [1, -1, \alpha^2, -\alpha^2, \alpha^{-2}, -\alpha^{-2}].$$

from which it follows that

$$[\mathrm{As}(V)] = [-1, 1, \alpha^2, \alpha^{-2}],$$
$$\mathrm{Ind}_H^G(\mathrm{ad}^0(V)) = [1, -1, \alpha^2, -\alpha^2, \alpha^{-2}, -\alpha^{-2}],$$

both of which have 1 as an eigenvalue. □

Lemma 7.5.18. *Assume* k *has characteristic* $p \geq 3$*. Let* G *be the group* $\mathrm{SL}_2(k) \wr \mathbf{Z}/2\mathbf{Z} = (\mathrm{SL}_2(k) \times \mathrm{SL}_2(k)) \rtimes \mathbf{Z}/2\mathbf{Z}$*, where the semi-direct product swaps the two copies of* $\mathrm{SL}_2(k)$*, considered as a subgroup of* $\mathrm{Sp}_4(k)$ *as in §2.2. Then* G *is weakly enormous, and is furthermore enormous if* $\#k \neq 5$*.*

Proof. We begin by checking that property (E1) holds. Let $H = \mathrm{SL}_2(k) \times \mathrm{SL}_2(k) = A \times B$, say, and let V_A and V_B denote the tautological 2-dimensional representations of A and B, so that $W|_H = V_A \oplus V_B$, and $\mathrm{Sym}^2(W)|_H = \mathrm{ad}^0(W)|_H = \mathrm{ad}^0(V_A) \oplus \mathrm{ad}^0(V_B) \oplus V_A \otimes V_B$. Since $H^1(G, \mathrm{ad}^0(W)) = H^1(H, \mathrm{ad}^0(W))^{G/H}$, it suffices to prove that

$$H^1(H, \mathrm{ad}^0(W)) = H^1(A \times B, \mathrm{ad}^0(W)) = 0.$$

By inflation–restriction, we see that there are exact sequences:

$$H^1(A, \mathrm{ad}^0(V_A)) \to H^1(A \times B, \mathrm{ad}^0(V_A)) \to (H^1(B, k) \otimes \mathrm{ad}^0(V_A))^A = 0,$$

$$0 = H^1(A, (V_A \otimes V_B)^B) \to H^1(A \times B, V_A \otimes V_B) \to (H^1(B, V_B) \otimes V_A)^A = 0.$$

Thus it remains to show that $H^1(A, \mathrm{ad}^0(V_A)) = 0$. But this is the same as showing that

$$H^1(\mathrm{SL}_2(k), \mathrm{Sym}^2(k^2)) = 0.$$

This holds for $\#k \neq 5$ (which we are assuming) by [DDT97, Lem. 2.48].

Property (E2) is obvious. For property (E3), it suffices by Lemma 7.5.17 to show that $G \setminus H$ contains an element g of order not dividing 4 and not divisible by p. Since $p^2 - 1$ is always divisible by 8, there exists a matrix $a \in A$ of order exactly 8. The automorphism $\sigma : A \times B \to B \times A$ of order 2 identifies A with B, and with respect to this identification let $g = \sigma(a, a) = (a, a)\sigma$. Then $g^2 = (a^2, a^2)$ has order 4, so g has order 8 which does not divide 4 and is not divisible by p, as required. □

For $p = 5$ one has the following substitute:

Lemma 7.5.19. *Let H/F be a quadratic extension, and let $\overline{r} : G_H \to \mathrm{GL}_2(\mathbf{F}_5)$ be a surjective representation with determinant $\overline{\varepsilon}^{-1}$. Let $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_5)$ be the induction of $\overline{r}$ to F, and assume that the image of $\overline{\rho}|_{G_{F(\zeta_5)}}$ is equal to $G = \mathrm{SL}_2(\mathbf{F}_5) \wr \mathbf{Z}/2\mathbf{Z}$. Assume furthermore that 5 is unramified in F. Then $G := \overline{\rho}(G_{F(\zeta_{5^N})})$ is weakly enormous for all $N \geq 1$ and ζ_5 does not lie in the fixed field of $\mathrm{ad}^0\,\overline{\rho}$; in particular, $\overline{\rho}$ is vast.*

Proof. Since the abelianization of G has order prime to 5, the image of $\overline{\rho}$ over $F(\zeta_5)$ is the same as the image over $F(\zeta_{5^N})$ for any N, and is weakly enormous by Lemma 7.5.18. Let Γ denote the image of $\overline{\rho}$. Since 5 is unramified in F and the similitude character of Γ is inverse cyclotomic, it follows that the similitude character is surjective and $[\Gamma : G] = 4$. In particular, the group Γ is the full pre-image of G in $\mathrm{GSp}_4(\mathbf{F}_p)$, and is generated by pairs (A, B) in $\mathrm{GL}_2(\mathbf{F}_p)$ with $\det(A) = \det(B)$ together with an involution sending (A, B) to (B, A). The fixed field L of $\mathrm{ad}^0\,\overline{\rho}$ is the fixed field of the projective representation. But one can now observe directly that the image of Γ in $\mathrm{PGSp}_4(\mathbf{F}_5)$ has abelianization $(\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z})$, which does not surject onto $\mathrm{Gal}(F(\zeta_5)/F) = \mathbf{Z}/4\mathbf{Z}$. So $\zeta_p \notin L$, and $\overline{\rho}$ is vast, as required. □

7.5.20. *The enormous subgroups of* $\mathrm{Sp}_4(\mathbf{F}_3)$. By an exhaustive search, one can determine precisely which of the subgroups of $\mathrm{Sp}_4(\mathbf{F}_3)$ are enormous. There are 162 conjugacy classes of subgroups, and it turns out that precisely 11 of them are enormous, of orders 40, 128, 160, 192, 240, 320, 384, 384, 1152, 1920, and 51840 respectively. Our main interest will be in representations $\overline{\rho}$ to $\mathrm{GSp}_4(\mathbf{F}_3)$ which are vast and tidy. In particular, it is of interest to consider subgroups G of $\mathrm{GSp}_4(\mathbf{F}_3)$ which are tidy and such that $H = \mathrm{Sp}_4(\mathbf{F}_3) \cap G$ is enormous. Sometimes the tautological 4-dimensional representation V of one of these groups G fails to be absolutely irreducible on an index two subgroup — necessarily this subgroup is not $H = G \cap \mathrm{Sp}_4(\mathbf{F}_3)$ because we are assuming that H is enormous and hence acts absolutely irreducibly on V. The representation V underlying G restricted to this index two subgroup either becomes reducible over $\mathbf{F}_3$ or over a non-trivial extension of $\mathbf{F}_3$. In the former case, we say that G is split induced. In this case, the index two subgroup is necessarily a subgroup of

$$\Delta = \{(A, B) \subset \mathrm{GL}_2(\mathbf{F}_p) \times \mathrm{GL}_2(\mathbf{F}_p) \text{ where } \det(A) = \det(B)\}$$

and G is a subgroup of $\Gamma := \Delta \rtimes \mathbf{Z}/2\mathbf{Z}$ where $\mathbf{Z}/2\mathbf{Z}$ swaps the factors. Hence we may write the index two subgroup in this case as $G \cap \Delta$.

We collect a number of interesting examples in the following lemma.

Lemma 7.5.21. *The following groups $G \subset \mathrm{GSp}_4(\mathbf{F}_3)$ are tidy, and such that the index two subgroup $H = G \cap \mathrm{Sp}_4(\mathbf{F}_3)$ is enormous.*

(1) The group $G = \mathrm{GSp}_4(\mathbf{F}_3)$.
(2) A group G of order 3840. The projective image has index 27 in $\mathrm{PGSp}_4(\mathbf{F}_3)$. It may be identified as the stabilizer of the natural action of $\mathrm{PGSp}_4(\mathbf{F}_3)$ on the 27 lines of a cubic surface.
(3) **Split Inductions**. The following subgroups of $\Gamma = \Delta \rtimes \mathbf{Z}/2\mathbf{Z}$:
   (a) The group $G = \Gamma$ of order 2304.
   (b) The two groups G of index 3 in Γ. They are the two groups of order 768 inside $\mathrm{GSp}_4(\mathbf{F}_3)$ up to conjugacy, and they are distinguished by their intersections $H = G \cap \mathrm{Sp}_4(\mathbf{F}_3) \subset \mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$ and $H \cap \Delta \subset \mathrm{SL}_2(\mathbf{F}_3)^2$. Note there is a homomorphism $\chi : \mathrm{SL}_2(\mathbf{F}_3) \to A_4 \to \mathbf{Z}/3\mathbf{Z}$. One intersection $H \cap \Delta$ is given by pairs (A, B) with $\chi(A) = \chi(B)$, and the other by pairs with $\chi(A) = -\chi(B)$. Note that these groups are abstractly isomorphic (the outer automorphism of $\mathrm{SL}_2(\mathbf{F}_3)$ sends χ to $-\chi$) but not conjugate inside $\mathrm{GSp}_4(\mathbf{F}_3)$.
(4) **Other Inductions.** A group G of order 480 with projective image $S_5 \times \mathbf{Z}/2\mathbf{Z}$. There is an isomorphism $\mathrm{PSL}_2(\mathbf{F}_9) = A_6$, and hence a projective $\mathbf{F}_9$ representation of the subgroup $A_5 \subset A_6$. This is not unique — there are two natural conjugacy classes of A_5 permuted by the exotic automorphism of A_6. But that automorphism is induced by Frob_3 acting on the field of coefficients $\mathbf{F}_9$, so the choice does not matter. There is a corresponding lift:

$$\widetilde{A}_5 \to \mathrm{GL}_2(\mathbf{F}_9)$$

by a group $\widetilde{A}_5$ which is a central extension of A_5 by $\mathbf{Z}/4\mathbf{Z}$. The outer automorphism group of $\widetilde{A}_5$ is $(\mathbf{Z}/2\mathbf{Z})^2$, and there is a unique such outer automorphism which acts by -1 on the centre and by an outer automorphism on A_5. Moreover, this lifts to a genuine automorphism σ of $\widetilde{A}_5$ of order 2. Then $G := \widetilde{A}_5 \rtimes \langle\sigma\rangle \subset \mathrm{GSp}_4(\mathbf{F}_3)$ has order 480. This is the only enormous subgroup which both has induced image and is not solvable. **Warning:** The group G is not determined up to conjugacy by its order. Indeed, there exists a second conjugacy class of subgroups G' of order 480 with $H' = G' \cap \mathrm{Sp}_4(\mathbf{F}_3)$ of order 240 such that G' contains $\widetilde{A}_5$ with index two and such that the corresponding outer automorphism is given by the class of σ. The group G', however, is not a semi-direct product. The groups G and G' can be distinguished as follows: the group $\mathrm{PGSp}_4(\mathbf{F}_3)$ has a natural action on 40 points corresponding to the action on $\mathbf{P}^3(\mathbf{F}_3)$. The orbits of G are of size 20 and 20 respectively whereas G' acts transitively.

Proof. This can be proved using the computer algebra package `magma` [BCGP21]. We omit the details. Note, however, that case (3a) was proved in Lemma 7.5.18. □

Lemma 7.5.22. *Suppose that $p \geq 3$, that K/F is a quadratic extension such that K is unramified at p, and that $\overline{r} : G_K \to \mathrm{GL}_2(k)$ restricted to $G_{K(\zeta_p)}$ has image $\mathrm{SL}_2(k)$. Choose $\sigma \in G_F \setminus G_K$, and assume that $\mathrm{Proj}\,\overline{r}^\sigma \not\cong \mathrm{Proj}\,\overline{r}$, but that*

$\det \overline{r}^\sigma = \det \overline{r}$ *is equal to* $\overline{\varepsilon}^{-1}$. *Let* $\overline{\rho} := \mathrm{Ind}_{G_K}^{G_F}\, \overline{r} : G_F \to \mathrm{GSp}_4(k)$. *If* $p = 3$, *assume that the fixed fields corresponding to the kernels of* $\mathrm{Proj}\,\overline{r}$ *and* $\mathrm{Proj}\,\overline{r}^\sigma$ *are disjoint. Then* $\overline{\rho}$ *is vast and tidy.*

Proof. Note that F is necessarily unramified at p (since K is). By Lemmas 7.5.18 and 7.5.19, in order to show that $\overline{\rho}$ is vast, it suffices to show that for all $N \geq 1$, the image of $\overline{\rho}$ restricted to $F(\zeta_{p^N})$ is $G = \mathrm{SL}_2(k) \wr \mathbf{Z}/2\mathbf{Z}$. First assume that $\#k > 3$. Then $\mathrm{PSL}_2(k)$ is simple. If the image of $\overline{r}^\sigma$ is disjoint from the image of $\overline{r}$, it would follow by Goursat's Lemma that the image of $\overline{\rho}|_{G_{K(\zeta_p)}}$ is the group $\mathrm{SL}_2(k)^2$, and hence the image of $\overline{\rho}|_{G_{K(\zeta_{p^N})}}$ is also $\mathrm{SL}_2(k)^2$, and thus the image of $\overline{\rho}|_{G_{F(\zeta_{p^N})}}$ is $\mathrm{SL}_2(k) \wr \mathbf{Z}/2\mathbf{Z}$. Since the automorphism group of $\mathrm{PSL}_2(k)$ is $\mathrm{PGL}_2(k)$, it follows that the projective representations associated to $\overline{r}$ and $\overline{r}^\sigma$ have the same image if and only if they are the same. Since we are assuming otherwise, we are done unless $k = \mathbf{F}_3$.

Now assume that $k = \mathbf{F}_3$, and so the images of $\overline{r}$ and $\overline{r}^\sigma$ restricted to $K(\zeta_3)$ are both isomorphic to $\mathrm{SL}_2(\mathbf{F}_3)$, which is a degree two central extension of A_4. The non-trivial quotients of $\mathrm{SL}_2(\mathbf{F}_3)$ are given by $\mathrm{PSL}_2(\mathbf{F}_3) = A_4$ and $\mathbf{Z}/3\mathbf{Z}$. By assumption, the fixed fields corresponding to the kernels of $\mathrm{Proj}\,\overline{r}$ and $\mathrm{Proj}\,\overline{r}^\sigma$ are disjoint and both have Galois group A_4. Thus by Goursat's lemma, the image of $\overline{\rho}$ restricted to $F(\zeta_3)$ is $\mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$. This is enormous, by Lemma 7.5.21. The abelianization of this group has order prime to 3, so the image of $\overline{\rho}$ restricted to $F(\zeta_{3^N})$ is also of this form.

Tidiness follows for $p \geq 5$ by Lemma 7.5.13. For $p = 3$, the image contains an element g of order 8 with $\nu(g) = -1$ and eigenvalues $(\zeta, -\zeta^{-1}, \zeta, -\zeta^{-1})$ for a primitive 8th root of unity ζ. The ratio of any two eigenvalues is either trivial or is a primitive fourth root of unity. □

Remark 7.5.23. When $p = 3$, we may weaken the hypotheses of this lemma slightly. By Lemma 7.5.21 and Lemma 7.5.5, it suffices that the image of $\overline{\rho}$ restricted to $F(\zeta_{27})$ is either $\mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$ or one of the subgroups of $\mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$ of index three and order 384 considered in Lemma 7.5.21. Unfortunately, the hypothesis that $\mathrm{Proj}\,\overline{r}$ is distinct from $\mathrm{Proj}\,\overline{r}^\sigma$ is not *quite* enough to force this. For example, it is possible that the image of $\overline{\rho}$ restricted to $F(\zeta_3)$ might be the (unique) subgroup of order 384 in $\mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$ with abelianization $\mathbf{Z}/6\mathbf{Z}$, and this means it is possible that the image of $\overline{\rho}$ restricted to $F(\zeta_9)$ is the 2-Sylow subgroup of $\mathrm{SL}_2(\mathbf{F}_3) \wr \mathbf{Z}/2\mathbf{Z}$ of order 128. However, this latter subgroup is not enormous.

7.5.24. *Crossing with dihedral extensions.* The goal of this section is to construct certain representations induced from quadratic fields K/F which will allow us to prove modularity results for elliptic curves over K even when K is neither totally real nor CM (see Theorem 10.1.4). Suppose that F is a totally real field in which p splits completely, and let K/F be an arbitrary quadratic extension of F in which p is unramified.

Lemma 7.5.25. *There exists a Galois extension* H/F *containing* K *such that:*

(1) $D = \mathrm{Gal}(H/F)$ *is the dihedral group of order* 8, *and* $\mathrm{Gal}(H/K) = (\mathbf{Z}/2\mathbf{Z})^2$.
(2) H/F *is the Galois closure over* F *of a quadratic extension* M/K.
(3) H/F *is unramified at each* $v|p$, *and* $\langle \mathrm{Frob}_v \rangle \in D$ *is not central.*

Furthermore, H/F *may be chosen to be linearly disjoint from any given fixed finite extension of* F *linearly disjoint from* K/F.

Proof. Let L/F be a second quadratic extension to be chosen later. The obstruction to constructing a dihedral extension H/F containing K and L as quadratic subfields with $\mathrm{Gal}(H/K) \simeq \mathrm{Gal}(H/L) \simeq (\mathbf{Z}/2\mathbf{Z})^2$ is the vanishing of the cup product $\chi_K \cup \chi_L$, where $\chi_K, \chi_L \in H^1(F, \mathbf{F}_2)$ are the quadratic characters corresponding to the fields K and L. Equivalently, if $L = F(\sqrt{\beta})$ and $K = F(\sqrt{\alpha})$, it is the condition of requiring that $\beta \in N_{K/F}(K^\times)$; if $\beta = N(x + y\sqrt{\alpha})$, then one may take

$$H = F(\sqrt{\alpha}, \sqrt{\beta}, \sqrt{x + y\sqrt{\alpha}}).$$

The extension $M = K(\sqrt{x + y\sqrt{\alpha}})$ will have Galois closure H over F. Suppose β can be chosen so that every $v|p$ is inert in L, and moreover such that β is prime to p. Then H/M will be unramified at each $v|p$, and Frob_v will be non-central, since the fixed field of the non-trivial central element is the compositum $K.L$.

We now construct many such β. Note that $F_v \simeq \mathbf{Q}_p$ by assumption, and we may assume that α is a v-adic unit for all $v|p$. Let us consider $N_{K/F}(K^\times)$, which consists of the non-zero elements of F of the form $x^2 - \alpha y^2$ where $x, y \in F$. The quadratic form

$$x^2 - \alpha y^2 - \gamma z^2 = 0 \mod v$$

for any $\gamma \neq 0$ always has a non-trivial solution with $z \neq 0$. Hence, by taking γ to be any quadratic non-residue in $\mathbf{F}_p^\times$, we may choose x and y modulo v so that $x^2 - \alpha y^2$ is a non-zero quadratic non-residue. Making such a choice for all $v|p$, we find that $\beta = x^2 - \alpha y^2$ is a v-adic unit and a quadratic non-residue modulo v for all $v|p$. Since $p > 2$, the resulting extension $L = F(\sqrt{\beta})$ is thus inert at all primes $v|p$, giving rise to the desired extension H/F.

Finally, by taking x and y sufficiently close to 1 and 0 respectively in $\mathcal{O}_{F,w}$ for any finite set of auxiliary primes w, can ensure that H/K splits completely at any such collection of primes, and hence we may ensure H/F is linearly disjoint from any fixed finite extension of F which is linearly disjoint from K, as required. □

We may write D as $D = \langle a, b | a^2 = b^2 = (ab)^4 = 1 \rangle$, where $[a, b]$ is the order two element of the centre of D.

Lemma 7.5.26. *Let $\overline{r} : G_F \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$ be an absolutely irreducible Galois representation with determinant $\overline{\varepsilon}^{-1}$. Suppose that, for each $v|p$, the restriction $\overline{r}|_{G_{F_v}}$ takes the shape*

$$\begin{pmatrix} \chi_v & * \\ 0 & \overline{\varepsilon}^{-1}\chi_v^{-1} \end{pmatrix}$$

for some unramified character χ_v. Let K/F be an arbitrary quadratic extension linearly disjoint from the fixed field $F(\overline{r})$ of the kernel of $\overline{r}$, let H/F be any corresponding D-extension as guaranteed by Lemma 7.5.25, chosen to be linearly disjoint from $F(\overline{r})$, and let M/K be a quadratic extension with Galois closure H/F. Let $\overline{\rho}$ be the following symplectic induction

$$\overline{\rho} := \mathrm{Ind}_{G_K}^{G_F}(\overline{r}|_{G_K} \otimes \delta_{M/K}),$$

where $\delta_{M/K}$ is the quadratic character corresponding to the extension M/K, and the induction is constructed as in §2.2. Let Γ denote the image of $\overline{r}$, and let G denote the image of $\overline{\rho}$. Then:

(1) $\overline{\rho}$ is weight 2 ordinary and p-distinguished with similitude character $\overline{\varepsilon}^{-1}$.

(2) *If V denotes the underlying representation of G given by $\overline{r}$, and U the 2-dimensional faithful representation of $\mathrm{Gal}(H/F) = D$, then $\overline{\rho}$ is given by $V \otimes U$. In particular, $\overline{\rho}$ is absolutely irreducible.*

(3) *If Γ has a central element of order 2, then the image of $\overline{\rho}$ is*

$$G = (\Gamma \times D)/(-1 = [a, b]).$$

Otherwise, the image is $G = \Gamma \times D$.

Proof. The restriction of $\overline{r}$ to G_K has determinant $\overline{\varepsilon}^{-1}$, which is preserved by the quadratic twist, and hence the induction also has $\overline{\varepsilon}^{-1}$ as the similitude character. The induction of $\delta_{M/K}$ from G_K to G_F is precisely the representation U of $D = \mathrm{Gal}(H/F)$.

By the construction of Lemma 7.5.25, for each place $v|p$, $\mathrm{Frob}_v \in D$ is not central. It follows that the restriction of $\mathrm{Gal}(H/F)$ acting on U to the decomposition group at v is of the form $\psi \oplus \chi$ for *distinct* unramified characters ψ and χ. Then the representation $\overline{\rho}|_{G_{F_v}}$ naturally takes the form $\overline{r}|_{G_{F_v}} \otimes \psi \oplus \overline{r}|_{G_{F_v}} \otimes \chi$. This is automatically weight 2 ordinary and p-distinguished.

Finally, the image of $\overline{\rho}$ is the image of the map $\Gamma \times D \to \mathrm{GL}(V \otimes U)$, and the kernel of this map is given by the elements of the form (z, z^{-1}) with z central. □

We now show that many of the groups G occurring as the image of representations $\overline{\rho}$ as constructed in Lemma 7.5.26 have big image.

Lemma 7.5.27. *Let $p \geq 5$, and suppose that we are in the setting of Lemma 7.5.26. Suppose either that $\Gamma = \mathrm{GL}_2(\mathbf{F}_p)$ or that $p = 5$ and Γ is the pre-image in $\mathrm{GL}_2(\mathbf{F}_5)$ of $S_4 \subset S_5 \cong \mathrm{PGL}_2(\mathbf{F}_5)$. Then $G \cap \mathrm{Sp}_4(\mathbf{F}_p)$ is enormous unless $\Gamma = \mathrm{GL}_2(\mathbf{F}_p)$ and $p = 5$, in which case $G \cap \mathrm{Sp}_4(\mathbf{F}_p)$ is weakly enormous. In any case, $\overline{\rho}$ is vast and tidy.*

Proof. By Lemma 7.5.26, G acts faithfully on $W = V \otimes U$, where V is the tautological representation of Γ, and U is the faithful 2-dimensional representation of D, with image $G = (\Gamma \times D)/(-1 = [a, b])$. We begin by checking condition (E3). Clearly

$$W \otimes W^* = (V \otimes V^*) \otimes (U \otimes U^*).$$

The latter factor is the regular representation of the abelianization of D, and is a direct sum of characters of order dividing 2. The first factor is the direct sum of $\mathrm{ad}^0(V)$ with the trivial character. Both the trivial representation and the adjoint representation of $\mathrm{GL}_2(\mathbf{F}_p)$ have the property that 1 is always an eigenvalue of any element. Hence, for any irreducible summand of $W \otimes W^*$, 1 will always be an eigenvalue on an index two subgroup $\Sigma \subset G$ which is the kernel of one of the degree 2 characters of D. Yet given $g \in \Gamma$, there is an element in Σ with eigenvalues the roots of g together with the negatives of the roots of g. Hence it suffices to note that $\Gamma \cap \mathrm{SL}_2(\mathbf{F}_p)$ has an element with eigenvalues $\{\alpha, \alpha^{-1}\}$ with $\alpha \neq \pm\alpha^{-1}$. (In particular, in the case that $p = 5$ and Γ is the central cover of S_4, one could take g to have order 3.)

For $\Gamma = \mathrm{GL}_2(\mathbf{F}_p)$, $p > 5$, since D has order prime to p, (E1) reduces to the fact that $H^1(\mathrm{SL}_2(\mathbf{F}_p), \mathrm{Sym}^2(\mathbf{F}_p^2)) = 0$, which is [DDT97, Lem. 2.48]. If $p = 5$, the group G is of order $384 = 4^2|S_4|$, and therefore satisfies (E1) automatically because the order is prime to p.

For the final claim, note firstly that for each $N \geq 1$, we have $\overline{\rho}(G_{F(\zeta_{p^N})}) = G \cap \mathrm{Sp}_4(\mathbf{F}_p)$. Indeed this is clear for $N = 1$ (as the similitude factor of $\overline{\rho}$ is ε^{-1}),

and since G has no quotients of order p, the same is true for all $N > 1$. That $\overline{\rho}$ is vast is then an immediate consequence of the previous claims except in the case when $\Gamma = \mathrm{GL}_2(\mathbf{F}_5)$, where $G \cap \mathrm{Sp}_4(\mathbf{F}_5)$ is not enormous. But in this case, exactly as in the proof of Lemma 7.5.19, the image of the projective representation factors through $\mathrm{PGL}_2(\mathbf{F}_5) \times (\mathbf{Z}/2\mathbf{Z})^2$ which does not surject onto $\mathbf{Z}/4\mathbf{Z}$, and hence the fixed field of the adjoint representation cannot contain ζ_5 when E is unramified at $p = 5$. Finally, for tidiness, we note that Γ and hence G contains a centre of order at least $p - 1$, and we are done by Lemma 7.5.12. □

7.6. **Taylor–Wiles primes.** We again fix a global deformation problem

$$\mathcal{S} = (\overline{\rho}, S, \{\Lambda_v\}_{v \in S}, \psi, \{\mathcal{D}_v\}_{v \in S}).$$

Then we define a *Taylor–Wiles datum* to be a tuple $(Q, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q})$ consisting of:

- A finite set of finite places Q of F, disjoint from S, such that $q_v \equiv 1 \bmod p$ for each $v \in Q$.
- For each $v \in Q$, an ordering $\overline{\alpha}_{v,1}, \overline{\alpha}_{v,2}, \overline{\alpha}_{v,3} = \overline{\psi}(\mathrm{Frob}_v)\overline{\alpha}_{v,2}^{-1}, \overline{\alpha}_{v,4} = \overline{\psi}(\mathrm{Frob}_v)\overline{\alpha}_{v,1}^{-1}$ of the eigenvalues of $\overline{\rho}(\mathrm{Frob}_v)$, which are assumed to be k-rational and pairwise distinct.

Given a Taylor–Wiles datum $(Q, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q})$, we define the augmented global deformation problem

$$\mathcal{S}_Q = (\overline{\rho}, S \cup Q, \{\Lambda_v\}_{v \in S} \cup \{\mathcal{O}\}_{v \in Q}, \psi, \{\mathcal{D}_v\}_{v \in S} \cup \{\mathcal{D}_v^{\square}\}_{v \in Q}).$$

Set $\Delta_Q = \prod_{v \in Q} \Delta_v$. For each $v \in Q$, the fixed ordering $\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4}$, determines a $\Lambda[\Delta_Q]$-algebra structure on $R^T_{\mathcal{S}_Q}$ for any subset T of S (via the homomorphisms $\mathcal{O}[\Delta_v] \to R_v^{\square}$ defined in §7.4.3). Letting $\mathfrak{a}_Q = \ker(\Lambda[\Delta_Q] \to \Lambda)$ be the augmentation ideal, the natural surjection $R^T_{\mathcal{S}_Q} \to R^T_{\mathcal{S}}$ has kernel $\mathfrak{a}_Q R^T_{\mathcal{S}_Q}$.

Lemma 7.6.1. *Assume that $\overline{\rho}$ is vast, that $p \geq 3$ is unramified in F, that $\psi = \varepsilon^{-1}$, and that k contains all of the eigenvalues of all elements of $\overline{\rho}(G_{F(\zeta_p)})$. Let $q \geq h^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1))$. Then for every $N \geq 1$, there is a choice of Taylor–Wiles datum $(Q_N, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q_N})$ satisfying the following:*

(1) $\#Q_N = q$.
(2) For each $v \in Q_N$, $q_v \equiv 1 \bmod p^N$.
(3) $h^1_{S^{\perp}_{Q_N}, S}(\mathrm{ad}^0\,\overline{\rho}(1)) = 0$.

Proof. Without loss of generality, we may assume that $N \geq 3$, and hence (by the definition of vastness and Remark 7.5.7) that $\overline{\rho}(G_{F(\zeta_{p^N})})$ is weakly enormous. By definition, we have

$$H^1_{S^{\perp}_{Q_N}, S}(\mathrm{ad}^0\,\overline{\rho}(1)) = \ker\left(H^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) \to \prod_{v \in Q_N} H^1(F_v, \mathrm{ad}^0\,\overline{\rho}(1))\right).$$

By induction, it suffices to show that given any cocycle κ representing a nonzero element of $H^1_{S^{\perp}_{Q_N}, S}(\mathrm{ad}^0\,\overline{\rho}(1))$, there are infinitely many finite places v of F such that

- v splits in $F(\zeta_{p^N})$;
- $\overline{\rho}(\mathrm{Frob}_v)$ has 4-distinct eigenvalues $\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4}$ in k;
- the image of κ in $H^1(F_v, \mathrm{ad}^0\,\overline{\rho}(1))$ is nonzero.

By Cebotarev, we are reduced to showing that given any cocycle κ representing a nonzero element of $H^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1))$, there is some $\sigma \in G_{F(\zeta_{p^N})}$ such that

- $\overline{\rho}(\sigma)$ has distinct (k-rational) eigenvalues;
- $p_\sigma \kappa(\sigma) \neq 0$, where $p_\sigma : \mathrm{ad}^0\,\overline{\rho} \to (\mathrm{ad}^0\,\overline{\rho})^\sigma$ is the σ-equivariant projection.

(The latter condition guarantees that the image of κ in $H^1(F_v, \mathrm{ad}^0\,\overline{\rho}(1))$ is not a coboundary.) Let L/F be the fixed field of $\mathrm{ad}^0\,\overline{\rho}$. The kernel of the restriction map

$$H^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) \to H^1(F_S/L(\zeta_{p^N}), \mathrm{ad}^0\,\overline{\rho}(1))^{G_F}$$

is, by inflation–restriction, isomorphic to

$$H^1(\mathrm{Gal}(L(\zeta_{p^N})/F), \mathrm{ad}^0\,\overline{\rho}(1)).$$

The assumption that $\overline{\rho}$ is vast implies by Lemma 7.5.9 that this group vanishes. In particular, the restriction of κ defines a nonzero $G_{F(\zeta_{p^N})}$-equivariant homomorphism $\mathrm{Gal}(F_S/L(\zeta_{p^N})) \to \mathrm{ad}^0\,\overline{\rho}$. Let W be a nonzero irreducible sub-$G_{F(\zeta_{p^N})}$-representation of the k-span of $\kappa(\mathrm{Gal}(F_S/L(\zeta_{p^N}))$. Since $\overline{\rho}(G_{F(\zeta_{p^N})})$ is weakly enormous and k is sufficiently large, there exists $\sigma_0 \in G_{F(\zeta_{p^N})}$ such that $\overline{\rho}(\sigma_0)$ has distinct k-rational eigenvalues and such that $W^{\sigma_0} \neq 0$ (this follows from the vastness assumption, in particular, by condition (E3) of 7.5.2). This implies that $\kappa(\mathrm{Gal}(F_S/L(\zeta_{p^N}))$ is not contained in the kernel of the σ_0-equivariant projection $p_{\sigma_0} : \mathrm{ad}^0\,\overline{\rho} \to (\mathrm{ad}^0\,\overline{\rho})^{\sigma_0}$. If $p_{\sigma_0}\kappa(\sigma_0) \neq 0$, then we take $\sigma = \sigma_0$. Otherwise, we choose $\tau \in G_{L(\zeta_{p^N})}$ such that $p_{\sigma_0}\kappa(\tau) \neq 0$, and we take $\sigma = \tau\sigma_0$; since $\overline{\rho}(\sigma) = \overline{\rho}(\sigma_0)$ and $\kappa(\sigma) = \kappa(\sigma_0) + \kappa(\tau)$, we are done. □

Definition 7.6.2. We say that $\overline{\rho} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{F}}_p)$ is *odd* if the similitude character ψ is odd, i.e. if for each place $v|\infty$ of F with corresponding complex conjugation c_v, we have $\psi(c_v) = -1$.

Corollary 7.6.3. *Assume that $\overline{\rho}$ is odd, that $\overline{\rho}$ is vast, and that k contains all of the eigenvalues of all elements of $\overline{\rho}(G_{F(\zeta_p)})$. Let $q \geq h^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1))$. Then for every $N \geq 1$, there is a choice of Taylor–Wiles datum $(Q_N, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q_N})$ satisfying the following:*

(1) $\#Q_N = q$.
(2) For each $v \in Q_N$, $q_v \equiv 1 \bmod p^N$.
(3) There is a local Λ-algebra surjection $R^{S,\mathrm{loc}}_{\mathcal{S}}[[X_1, \ldots, X_g]] \to R^S_{\mathcal{S}_{Q_N}}$ with $g = 2q - 4[F : \mathbf{Q}] + \#S - 1$.

Proof. By Proposition 7.2.1 and Theorem 7.6.1, the claim holds with g instead equal to

$$\#S - 1 - \sum_{v|\infty} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}) + \sum_{v \in Q_N} h^0(F_v, \mathrm{ad}^0\,\overline{\rho}(1)).$$

(Note that the assumption that $\overline{\rho}$ is vast implies that $h^0(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)) = 0$.) For $v \in Q_N$, by the assumptions that $q_v \equiv 1 \bmod p$ and that $\overline{\rho}|_{G_{F_v}}$ has distinct eigenvalues we have

$$h^0(F_v, \mathrm{ad}^0\,\overline{\rho}(1)) = h^0(F_v, \mathrm{ad}^0\,\overline{\rho}) = 2.$$

For $v|\infty$ we have $h^0(F_v, \mathrm{ad}^0\,\overline{\rho}) = 4$ by the assumption that $\overline{\rho}$ is odd. It follows that $g = 2q - 4[F : \mathbf{Q}] + \#S - 1$, as claimed. □

7.7. **Global Galois deformation problems.** We now begin to introduce the framework that we need to carry out our Taylor–Wiles patching argument. As always, F is a totally real field in which the prime $p \geq 3$ splits completely, and we write S_p for the set of primes of F dividing p. Let $\overline{\rho} := G_F \to \mathrm{GSp}_4(k)$ be an absolutely irreducible representation. We assume the following hypotheses.

Hypothesis 7.7.1.

(1) The representation $\overline{\rho}$ is vast and tidy.
(2) If $v \in S_p$, then $\overline{\rho}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary.
(3) There is a set of finite places R of F which is disjoint from S_p, such that
  (a) If $v \in R$, then $\overline{\rho}|_{G_{F_v}}$ is trivial, and $q_v \equiv 1 \pmod{p}$. If $p = 3$ then we further insist that $q_v \equiv 1 \pmod{9}$.
  (b) If $v \notin S_p \cup R$, then $\overline{\rho}|_{G_{F_v}}$ is unramified.

Set $\psi = \varepsilon^{-1}$, and drop ψ from our notation for global deformation problems from now on. Let $I \subset S_p$ be a set of places of cardinality $\#I$. We will eventually need to assume that $\#I \leq 1$, although the more formal parts of the patching construction can be carried out without this assumption, so we do not impose it yet. We write I^c for $S_p \setminus I$.

By the Cebotarev density theorem and our assumption that $\overline{\rho}(G_F)$ is tidy, we can find an unramified place $v_0 \notin R \cup S_p$ of F with the properties that

- $q_{v_0} \not\equiv 1 \pmod{p}$,
- no two eigenvalues of $\overline{\rho}(\mathrm{Frob}_{v_0})$ have ratio q_{v_0}, and
- v_0 has residue characteristic greater than 5.

Then $H^2(F_{v_0}, \mathrm{ad}\,\overline{\rho}) = H^0(F_{v_0}, \mathrm{ad}\,\overline{\rho}(1))^\vee = 0$. We set $S = R \cup S_p \cup \{v_0\}$.

The reason for choosing v_0 is that all liftings of $\overline{\rho}|_{G_{F_{v_0}}}$ are automatically unramified by Proposition 7.4.2, and our choice of level structure at v_0 will guarantee that our level structures will be neat, by Lemma 7.8.3.

For each $v \in R$ we choose a pair of characters $\chi_v = (\chi_{v,1}, \chi_{v,2})$, where $\chi_{v,i} : \mathcal{O}_{F_v}^\times \to \mathcal{O}^\times$ are trivial modulo λ. (Note that at this stage the characters $\chi_{v,i}$ are allowed to be trivial.) We write χ for the tuple $(\chi_v)_{v \in R}$ as well as for the induced character $\chi = \prod_{v \in R} \chi_v : \prod_{v \in R} \mathrm{Iw}(v) \to \mathcal{O}^\times$.

For each place $v|p$, we fix Λ_v, (and thus θ_v) as in §7.3, in the following way: if $v \in I$, then we take $\Lambda_v = \mathcal{O}[[\mathcal{O}_{F_v}^\times(p)]]$, while if $v \notin I$, then we take $\Lambda_v = \mathcal{O}[[(\mathcal{O}_{F_v}^\times(p))^2]]$. We write $\mathfrak{r} = \{\overline{\mathfrak{r}}_v\}_{v \in S_p}$ for a choice of $\overline{\alpha}_v$ or $\overline{\beta}_v$ at each $v \in S_p$.

We have the corresponding global deformation problem

$$\mathcal{S}_\chi^{I,\mathfrak{r}} = (\overline{\rho}, S, \{\Lambda_{v,1}\}_{v \in I} \cup \{\Lambda_{v,2}\}_{v \in I^c} \cup \{\mathcal{O}\}_{v \in S \setminus S_p}, \\ \{\mathcal{D}_v^P\}_{v \in I} \cup \{\mathcal{D}_v^{B,\overline{\mathfrak{r}}_v}\}_{v \in I^c} \cup \{\mathcal{D}_v^\chi\}_{v \in R} \cup \{\mathcal{D}_{v_0}^\square\}).$$

Let $(Q, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,n})_{v \in Q})$ be a choice of Taylor–Wiles datum. We set $S_Q = S \cup Q$ and define the associated global deformation problem

$$\mathcal{S}_{\chi,Q}^{I,\mathfrak{r}} = (\overline{\rho}, S_Q, \{\Lambda_{v,1}\}_{v \in I} \cup \{\Lambda_{v,2}\}_{v \in I^c} \cup \{\mathcal{O}\}_{v \in S \setminus S_p}, \\ \{\mathcal{D}_v^P\}_{v \in I} \cup \{\mathcal{D}_v^{B,\overline{\mathfrak{r}}_v}\}_{v \in I^c} \cup \{\mathcal{D}_v^\chi\}_{v \in R} \cup \{\mathcal{D}_v^\square\}_{v \in S_Q \setminus (R \cup S_p)}).$$

Note that by definition $\mathcal{S}_{\chi,Q}^{I,\mathfrak{r}}$ does not depend on the choice of $\mathfrak{r}_v$ for $v \in I$.

7.8. **Taylor–Wiles systems: initial construction.** In the next two sections, we will construct the Taylor–Wiles systems that we will patch in §7.11, using an abstract patching criterion explained in §7.10.1. (§7.8 is mainly concerned with the construction of the Taylor–Wiles systems, whereas §7.9 is mainly concerned with proving the required local–global compatibility statements for the corresponding Galois representations.)

Since we are only dealing with the cases that $\#I \leq 1$, we do not need to make use of the full machinery of patching complexes developed in [CG18, KT17, GN20]; rather, we can and do use the notion of "balanced" modules introduced in [CG18, §2], which we recalled in §2.10. This has the advantage that we do not need to consider local global compatibility at places dividing p for Galois representations associated to classes in higher degrees of cohomology, but rather just have to prove the vanishing of the Euler characteristic of a certain perfect complex, which follows from a calculation of the cohomology in terms of automorphic forms.

We now make the following hypotheses on a representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(k)$, which include those made in Hypothesis 7.7.1.

Hypothesis 7.8.1.

(1) F is a totally real field in which the prime $p \geq 3$ splits completely; we write S_p for the set of primes of F dividing p.
(2) The representation $\overline{\rho}$ is vast and tidy.
(3) For each $v \in S_p$, $\overline{\rho}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary.
(4) There is a set of finite places R of F which is disjoint from S_p, such that
   (a) If $v \in R$, then $\overline{\rho}|_{G_{F_v}}$ is trivial, and $q_v \equiv 1 \pmod{p}$. If $p = 3$, then $q_v \equiv 1 \pmod{9}$.
   (b) If $v \notin S_p \cup R$, then moreover $\overline{\rho}|_{G_{F_v}}$ is unramified.
(5) There is an ordinary cuspidal automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 with central character $|\cdot|^2$ such that:
   (a) $\overline{\rho}_{\pi,p} \cong \overline{\rho}$.
   (b) If $v \in R \cup S_p$, then $\pi_v^{\mathrm{Iw}(v)} \neq 0$.
   (c) If $v \notin R \cup S_p$, then $\pi_v^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \neq 0$.

As in §7.7, by the assumption that $\overline{\rho}(G_F)$ is tidy we can and do choose an unramified place $v_0 \notin R \cup S_p$ with the properties that

- $q_{v_0} \not\equiv 1 \pmod{p}$,
- no two eigenvalues of $\overline{\rho}(\mathrm{Frob}_{v_0})$ have ratio q_{v_0}, and
- the residue characteristic of v_0 is greater than 5.

Definition 7.8.2. We define an open compact subgroup $K^p = \prod_v K_v$ of $\mathrm{GSp}_4(\mathbf{A}_F^{\infty,p})$ as follows:

- If $v \notin S_p \cup R \cup \{v_0\}$, then $K_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$.
- If $v \in R \cup \{v_0\}$, then $K_v = \mathrm{Iw}_1(v)$.

For any Taylor–Wiles datum $(Q, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q})$, we have open compact subgroups $K_0^p(Q)$, $K_1^p(Q)$ of K^p given by

- If $v \notin Q$, then $K_0^p(Q)_v = K_1^p(Q)_v = K_v^p$.
- If $v \in Q$, then $K_0^p(Q)_v = \mathrm{Iw}(v)$, $K_1^p(Q)_v = \mathrm{Iw}_1(v)$.

We define the open compact subgroup group $K_0^p(Q, R)$ as follows:

- If $v \notin Q \cup R$, then $K_0^p(Q, R)_v = K_v^p$.

- If $v \in Q \cup R$, then $K_0^p(Q,R)_v = \mathrm{Iw}(v)$.

Finally, we let $K_1^p(Q,R) = K_1^p(Q)$. (Note that we already have $K_1^p(Q)_v = \mathrm{Iw}_1(v)$ for $v \in R$.)

The following lemma (applied with $v = v_0$) guarantees that for any compact open subgroup $K_p \subset \mathrm{GSp}_4(F_p)$, $K_pK_0^p(Q)$ and $K_pK_1^p(Q)$ are neat.

Lemma 7.8.3. *Suppose that $K = \prod_v K_v \subset \mathrm{GSp}_4(\mathbf{A}_F^\infty)$ is an open compact subgroup and that there exists a place v of F such that v is absolutely unramified of residue characteristic greater than 5, and $K_v = \mathrm{Iw}_1(v)$. Then K is neat.*

Proof. Suppose that there is an element $g_v \in K_v$ which has an eigenvalue $\zeta \in \overline{F}_v$ which is a root of unity; by the definition of "neat" (see Definition 3.2.1), it is enough to check that we must have $\zeta = 1$. Since the reduction modulo v of the characteristic polynomial of g is $(X-1)^4$, the v-adic valuation of $(1-\zeta)$ is at least $1/4$. On the other hand, if v has residue characteristic l and $\zeta \neq 1$ is a root of unity, then the v-adic valuation of $(1-\zeta)$ is either 0, or is at most $1/(l-1)$, so we are done, as $l > 5$ by assumption. □

We let

$$\widetilde{\mathbf{T}} = \bigotimes_{v \notin S_p \cup R \cup \{v_0\}} \mathcal{O}[\mathrm{GSp}_4(F_v) /\!/ \mathrm{GSp}_4(\mathcal{O}_{F_v})]$$

be the ring of spherical Hecke operators away from the bad places, and similarly we set

$$\widetilde{\mathbf{T}}^Q = \bigotimes_{v \notin S_p \cup R \cup \{v_0\} \cup Q} \mathcal{O}[\mathrm{GSp}_4(F_v) /\!/ \mathrm{GSp}_4(\mathcal{O}_{F_v})].$$

We let $\widetilde{\mathfrak{m}}^{\mathrm{an}} \subset \widetilde{\mathbf{T}}$ be the maximal ideal corresponding to $\overline{\rho}$ (the "an" stands for "anaemic"); so by definition $\mathfrak{m}$ containins λ, and the polynomials $\det(X - \overline{\rho}(\mathrm{Frob}_v))$ and $Q_v(X)$ are congruent modulo $\mathfrak{m}$ for each $v \notin S_p \cup R \cup \{v_0\}$, where in a slight abuse of notation, if $v \notin S_p \cup R \cup \{v_0\}$ we write $Q_v(X) \in \widetilde{\mathbf{T}}[X]$ for the polynomial

$$X^4 - T_{v,1}X^3 + (q_vT_{v,2} + (q_v^3 + q_v)T_{v,0})X^2 - q_v^3T_{v,0}T_{v,1}X + q_v^6T_{v,0}^2$$

(cf. (2.4.8)). Similarly we write $\widetilde{\mathfrak{m}}^{\mathrm{an},Q} \subset \widetilde{\mathbf{T}}^Q$ for the maximal ideal corresponding to $\overline{\rho}$. For any choice of I we let

$$\widetilde{\mathbf{T}}^I = \widetilde{\mathbf{T}}[\{U_{v,0}, U_{\mathrm{Kli}(v),1}, U_{v,2}\}_{v \in I}, \{U_{v,0}, U_{v,1}, U_{v,2}\}_{v \in I^c}]$$

and

$$\widetilde{\mathbf{T}}^{I,Q} = \widetilde{\mathbf{T}}^Q[\{U_{v,0}, U_{\mathrm{Kli}(v),1}, U_{v,2}\}_{v \in I}, \{U_{v,0}, U_{v,1}, U_{v,2}\}_{v \in I^c}]$$

and additionally for any choice of $\mathfrak{s}$ we let $\widetilde{\mathfrak{m}}^{I,\mathfrak{s}} \subset \widetilde{\mathbf{T}}^I$ be the maximal ideal

(7.8.4)

$$\widetilde{\mathfrak{m}}^{I,\mathfrak{s}} = (\widetilde{\mathfrak{m}}^{\mathrm{an}}, \{U_{v,0} - 1, U_{v,2} - \alpha_v\beta_v\}_{v \in S_p}, \{U_{\mathrm{Kli}(v),1} - \alpha_v - \beta_v\}_{v \in I}, \{U_{v,1} - \mathfrak{s}_v\}_{v \in I^c})$$

and we let $\widetilde{\mathfrak{m}}^{I,\mathfrak{s},Q} \subset \widetilde{\mathbf{T}}^{I,Q}$ be the maximal ideal

$$\widetilde{\mathfrak{m}}^{I,\mathfrak{s},Q} = (\widetilde{\mathfrak{m}}^{\mathrm{an},Q}, \{U_{v,0} - 1, U_{v,2} - \alpha_v\beta_v\}_{v \in S_p}, \{U_{\mathrm{Kli}(v),1} - \alpha_v - \beta_v\}_{v \in I}, \{U_{v,1} - \mathfrak{s}_v\}_{v \in I^c}).$$

Let $\chi = (\chi_{v,1}, \chi_{v,2})_{v \in R}$ be any choice of p-power order characters of I_{F_v} for $v \in R$, and also write χ_v for the corresponding characters of $T(k(v))$ given by $\chi_{v,1} \circ \mathrm{Art}_{F_v}$, $\chi_{v,2} \circ \mathrm{Art}_{F_v}$.

Then we consider the Λ_I-module

$$M^{\chi,I,\mathfrak{s}} = \mathrm{RHom}^0_{\Lambda_I}(M^{\bullet,I}_{K^p}, \Lambda_I)_{\widetilde{\mathfrak{m}}^{I,\mathfrak{s}},\chi,|\cdot|^2},$$

and the $\Lambda_I[\Delta_Q]$-module

$$M^{\chi,I,\mathfrak{f},Q} = \mathrm{RHom}^0_{\Lambda_I}(M^{\bullet,I}_{K_1^p(Q)}, \Lambda_I)_{\widetilde{\mathfrak{m}}^{I,\mathfrak{f},Q},\widetilde{\mathfrak{m}}_Q,\chi,|\cdot|^2},$$

where:

- $M^{\bullet,I}_{K^p}$ denotes the complex $M^\bullet_I$ defined in Theorem 4.6.1, at tame level K^p.
- The localizations $\widetilde{\mathfrak{m}}^{I,\mathfrak{f}}, \widetilde{\mathfrak{m}}^{I,\mathfrak{f},Q}$ are defined above.
- The localization $\widetilde{\mathfrak{m}}_Q$ is with respect to the maximal ideals $\widetilde{\mathfrak{m}}_v$ of the subalgebras $\mathcal{O}[T(F_v)/T(\mathcal{O}_{F_v})_1]$ of the pro-v Iwahori Hecke algebras $\mathcal{H}_1(v)$ for $v \in Q$ as considered in §2.4.29, so that $\lambda \in \widetilde{\mathfrak{m}}_v$, $U_{v,0} - 1 \in \widetilde{\mathfrak{m}}_v$, and $U_{v,1}$ and $U_{v,2}$ are respectively congruent to $\overline{\alpha}_{v,1}$, $\overline{\alpha}_{v,1}\overline{\alpha}_{v,2}$ modulo $\widetilde{\mathfrak{m}}_v$.
- The subscript χ denotes that we take the χ-coinvariants for the action of $\prod_{v\in R} T(k(v))$.
- The subscript $|\cdot|^2$ denotes that we are fixing the central character, by taking coinvariants under $T_{v,0} - q_v^{-2}$ for all $v \notin S_p \cup R \cup \{v_0\}$.

The following lemma motivates our definition using $\mathrm{RHom}^0_\Lambda(M^\bullet, \Lambda)$, and will be useful for proving various properties of $M^{\chi,I,\mathfrak{f},W}$ below (see also Remark 7.8.7).

Lemma 7.8.5. *Let $\Lambda \in \mathrm{CNL}_\mathcal{O}$, and let $M^\bullet$ be a perfect complex of Λ-modules bounded below by 0. Set $M := \mathrm{RHom}^0_\Lambda(M^\bullet, \Lambda)$. Then, writing $*$ for the usual duality of finite-dimensional vector spaces and $^\vee$ for Pontryagin duals, we have*

(1) $M \otimes_\Lambda k = (H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda k))^*$.
(2) For any homomorphism of $\mathcal{O}$-algebras $\Lambda \to E$, $M \otimes_\Lambda E = (H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda E))^*$.
(3) For any homomorphism of $\mathcal{O}$-algebras $\Lambda \to \mathcal{O}$,

$$M \otimes_\Lambda \mathcal{O} = \mathrm{Hom}(H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda E/\mathcal{O}), E/\mathcal{O}) = H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda E/\mathcal{O})^\vee.$$

Proof. Let $P^\bullet = P^0 \to P^1 \to \ldots \to P^{l_0}$ be a bounded complex of finite projective Λ-modules which is bounded below by 0 and is quasi-isomorphic to $M^\bullet$. Then, by definition, we have an exact sequence

$$\mathrm{Hom}_\Lambda(P^1, \Lambda) \to \mathrm{Hom}_\Lambda(P^0, \Lambda) \to M \to 0.$$

In particular it follows that for any Λ-algebra R, we have an exact sequence of R-modules

$$\mathrm{Hom}_R(P^1 \otimes_\Lambda R, R) \to \mathrm{Hom}_R(P^0 \otimes_\Lambda R, R) \to M \otimes_\Lambda R \to 0.$$

On the other hand, by definition, we have an exact sequence of R-modules

$$0 \to H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda R) \to P^0 \otimes_\Lambda R \to P^1 \otimes_\Lambda R,$$

and therefore, for a field $R = F$, an exact sequence

$$\mathrm{Hom}_F(P^1 \otimes_\Lambda F, F) \to \mathrm{Hom}_F(P^0 \otimes_\Lambda F, F) \to \mathrm{Hom}_F(H^0(M^\bullet \otimes^{\mathbf{L}}_\Lambda F), F) \to 0.$$

Parts (1) and (2) follow immediately with $F = E$ or $F = k$. Part (3) follows from Lemma 7.8.6 below, applied to the morphism $P^0 \otimes_\Lambda \mathcal{O} \to P^1 \otimes_\Lambda \mathcal{O}$. □

Lemma 7.8.6. *If $\phi : M \to N$ is a morphism of finite free $\mathcal{O}$-modules, and $\phi_{E/\mathcal{O}} = \phi \otimes E/\mathcal{O}$ is the map $M \otimes E/\mathcal{O} \to N \otimes E/\mathcal{O}$, then the Pontryagin dual $\phi^\vee_{E/\mathcal{O}}$ of $\phi_{E/\mathcal{O}}$ is the map*

$$\phi^\vee_{E/\mathcal{O}} : \mathrm{Hom}(N, \mathcal{O}) \to \mathrm{Hom}(M, \mathcal{O}).$$

In particular, the Pontryagin dual of $\ker(\phi_{E/\mathcal{O}})$ is $\mathrm{coker}(\phi^\vee_{E/\mathcal{O}})$.

Proof. Because M and N are free, the Pontryagin duals of $M \otimes E/\mathcal{O}$ and $N \otimes E/\mathcal{O}$ are $\mathrm{Hom}(M, \mathcal{O})$ and $\mathrm{Hom}(N, \mathcal{O})$ respectively, and the result follows immediately. □

Remark 7.8.7. In [CG18] and [CG20], the patched modules are constructed by first taking cohomology with coefficients in $E/\mathcal{O}$ and then taking Pontryagin duals. Lemma 7.8.5 (3) explains how our construction coincides with this in the special case when $\Lambda = \mathcal{O}$.

Definition 7.8.8. For any $I \subset S_p$, a *weight* is a homomorphism $\kappa : \Lambda_I \to \mathcal{O}$; by definition, κ corresponds to a tuple $(\theta_{v,1}, \theta_{v,2})_{v \in S_p}$ where $\theta_{v,i} : I_{F_v} \to \mathcal{O}^\times$ is a character with trivial reduction, and moreover $\theta_{v,1} = \theta_{v,2}$ for $v \in I$. We let $\mathfrak{p}_\kappa \subset \Lambda_I$ denote the kernel of this homomorphism.

We say that κ is *classical* if there are integers $k_v \geq l_v \geq 2$ such that $\theta_{v,1} = \varepsilon^{(k_v+l_v)/2-2}$, $\theta_{v,2} = \varepsilon^{(k_v-l_v)/2}$ (so that $k_v \equiv l_v \equiv 2$ or $p+1 \pmod{2(p-1)}$, and if $v \in I$, we must have $l_v = 2$). If κ is classical, then we write ω^κ for the automorphic vector bundle corresponding to $(k_v, l_v)_{v \in S_p}$, as in §3.7.

For any I we denote by κ_2 the classical algebraic weight where $k_v = l_v = 2$ for all v. For $I = \emptyset$ we pick some sufficiently regular classical algebraic weight, κ_{reg}; for example, we could choose the one given by the characters $\theta_{v,1} = \varepsilon^{2N(p-1)}$ and $\theta_{v,2} = \varepsilon^{N(p-1)}$ for all $v \in S_p$, where N is sufficiently large.

Remark 7.8.9. In practice we choose κ_{reg} so that we can apply Theorems 3.10.1 and 6.6.5 in weight κ_{reg}. We will do this without comment from now on.

We will now prove some very important properties of the action of Δ_Q on the modules that we patch. It will also be important for us to understand the action of the diamond operators at the places in R (that is, at the places involved in the "Ihara avoidance" argument). We can and do treat the places in Q and in R simultaneously; recall, by Definition 7.8.2, we have the groups $K_1^p(Q)$ and $K_0^p(Q, R)$ such that

- If $v \notin Q \cup R$, then $K_0^p(Q, R)_v = K_v^p$ which equals $K_1^p(Q)_v$.
- If $v \in Q \cup R$, then $K_0^p(Q, R)_v = \mathrm{Iw}(v)$ which contains $K_1^p(Q)_v = \mathrm{Iw}_1(v)$.

In particular, there is an inclusion $K_1^p(Q) \subset K_0^p(Q, R)$. In contexts in which we particularly want to emphasize the fact that $K_1^p(Q)$ has level structure $\mathrm{Iw}_1(v)$ at $v \in R$, we write $K_1^p(Q, R) = K_1^p(Q)$.

Let K_p be any reasonable level structure at p (for example $K_p(I)$). Let $X_{K_p K_0^p(Q,R),\Sigma}$ be the Shimura variety of the corresponding level $K_p K_0^p(Q, R)$ for a choice Σ of good polyhedral cone decomposition. Over the interior $Y_{K_p K_0^p(Q,R)}$ we have for all $v \in Q \cup R$ a flag of subgroups $0 \subset H_v \subset L_v \subset H_v^\perp \subset A[v]$ and all the graded pieces are étale $k(v)$-group schemes of rank 1. We now consider the Shimura variety $X_{K_p K_1^p(Q,R),\Sigma}$ for the same choice of cone decomposition.

Proposition 7.8.10.

(1) *For all $v \in Q \cup R$, the groups H_v, L_v/H_v, $H_v^\perp/L_v$ and $A[v]/(H_v^\perp)$ extend to finite étale $k(v)$-group schemes of rank 1 over $X_{K_p K_0^p(Q,R),\Sigma}$.*

(2) *The map $X_{K_p K_1^p(Q,R),\Sigma} \to X_{K_p K_0^p(Q,R),\Sigma}$ is finite étale with group $\prod_{v \in Q \cup R} T(k(v))$, and $X_{K_p K_1^p(Q,R),\Sigma}$ identifies with the torsor of trivializations of the groups H_v, L_v/H_v, $H_v^\perp/L_v$ and $A[v]/(H_v^\perp)$, compatible with duality.*

Proof. We observe that when $F = \mathbf{Q}$, this is the content of [Str15, §2.4.5]. The argument can be adapted to our setting. The extension problem is local so let us pick $\sigma \in \Sigma$ and consider the completion $(X_{K_pK_0^p(Q,R),\Sigma})^\wedge_\sigma \simeq \mathrm{Spf}\ R$ of $X_{K_pK_0^p(Q,R),\Sigma}$ along the σ-stratum. The semi-abelian scheme A over $\mathrm{Spf}\ R$ is obtained by Mumford's construction as the quotient of a semi-abelian scheme B of constant toric rank by a finite free $\mathcal{O}_F$-module X_σ. Let $U_\sigma \hookrightarrow \mathrm{Spec}\, R$ be the Zariski open complement of the boundary and let us consider any of the groups H_v, L_v/H_v, $H_v^\perp/L_v$ or $A[v]/(H_v^\perp)$. If this group is a subquotient of $B[v]$, then since B exists over all $\mathrm{Spec}\, R$ and $B[v]$ is a finite étale group scheme, the group extends as a subquotient of $B[v]$. Otherwise, the group maps isomorphically to its image in $A[v]/B[v] = X_\sigma \otimes_{\mathcal{O}_F} k(v)$ and is constant over U_σ. Therefore it extends to the constant group scheme. This proves (1).

We may now define a scheme $X'_{K_pK_1^p(Q,R),\Sigma} \to X_{K_pK_0^p(Q,R),\Sigma}$ as the torsor of trivializations of the (extended) groups H_v, L_v/H_v, $H_v^\perp/L_v$ and $A[v]/(H_v^\perp)$, compatible with duality for all $v|p$. This scheme is canonically isomorphic to $X_{K_pK_1^p(Q,R),\Sigma}$ because the two schemes are generically equal, and both are normal, and finite flat over $X_{K_pK_0^p(Q,R),\Sigma}$. □

Proposition 7.8.11.

(1) $M^{\chi,\emptyset,\mathfrak{s},Q}$ is a finite free $\Lambda_\emptyset[\Delta_Q]$-module.
(2) If $\#I = 1$, then $M^{\chi,I,\mathfrak{s},Q}$ is a balanced $\Lambda_I[\Delta_Q]$-module.

Proof. The complex $M^{\bullet,I}_{K_1^p(Q)}$ (which is the complex $M_I^\bullet$ defined in Theorem 4.6.1 for the tame level $K_1^p(Q)$) is a perfect complex of Λ_I-modules of amplitude $[0, \#I]$. We claim that it is actually a perfect complex of $\Lambda_I[\prod_{v\in Q\cup R} T(k(v))]$-modules of amplitude $[0, \#I]$.

The complex $M^{\bullet,I}_{K_1^p(Q)}$ is obtained by considering the cohomology over $\mathfrak{X}_{K_pK_1^p(Q),\mathrm{Kli}(p^\infty)}$ of the sheaf of Λ_I-modules $\Omega^{\kappa_I}(-D)$ and applying the ordinary idempotent. Equivalently, it is obtained by considering the cohomology over $\mathfrak{X}_{K_pK_0^p(Q),\mathrm{Kli}(p^\infty)}$ of the sheaf of Λ_I-modules

$$\Omega^{\kappa_I}(-D) \otimes_{\mathcal{O}_{\mathfrak{X}_{K_pK_0^p(Q,R),\mathrm{Kli}(p^\infty)}}} \mathcal{O}_{\mathfrak{X}_{K_pK_1^p(Q,R),\mathrm{Kli}(p^\infty)}}$$

and applying the ordinary idempotent. Using the independence of the cohomology with respect to choices of toroidal compactifications, we may assume that we are in the setting of Proposition 7.8.10, so that the morphism $\mathfrak{X}_{K_pK_1^p(Q,R),\mathrm{Kli}(p^\infty)} \to \mathfrak{X}_{K_pK_0^p(Q,R),\mathrm{Kli}(p^\infty)}$ is finite étale with group $\prod_{v\in Q\cup R} T(k(v))$. Therefore, it follows (by considering a suitable étale covering to compute the cohomology) that $M^{\bullet,I}_{K_1^p(Q)}$ is represented by a bounded complex of flat complete $\Lambda_I[\prod_{v\in Q\cup R} T(k(v))]$-modules. We can apply Lemma 4.6.22 (or rather its straightforward extension to the semi-local situation; see also [Nak84, Prop. 2]) to conclude that $M^{\bullet,I}_{K_1^p(Q)}$ is a perfect complex of $\Lambda_I[\prod_{v\in Q\cup R} T(k(v))]$-modules of amplitude $[0, \#I]$. It follows that the corresponding complex $M^{\bullet,I,\chi}_{K_1^p(Q)}$ is a perfect complex of $\Lambda_I[\Delta_Q]$-modules, also of amplitude $[0, \#I]$.

Given an ideal $\widetilde{\mathfrak{m}}^{I,\mathfrak{s}}$, we can f localize the complex with respect to the action of a lift of a suitable idempotent for this ideal in the Hecke algebra, and this localization also preserves the property of being perfect of the correct amplitude. (The endomorphism ring at the level of derived categories of a perfect complex

of Λ_I modules is a finite Λ_I module. So, if one has a commutative subalgebra, it is a semi-local ring. See the discussion following [KT17, Lem. 2.12] for a lengthier treatment of such localizations.)

It remains to consider the passage to coinvariants under the centre. To this end, consider the spaces

$$\widetilde{M}^{\chi,I,\mathfrak{r}} = \mathrm{RHom}^0_{\Lambda_I}(M^{\bullet,I}_{K^p},\Lambda_I)_{\widetilde{\mathfrak{m}}^{I,\mathfrak{r}},\chi},$$

$$\widetilde{M}^{\chi,I,\mathfrak{r},Q} = \mathrm{RHom}^0_{\Lambda_I}(M^{\bullet,I}_{K^p_1(Q)},\Lambda_I)_{\widetilde{\mathfrak{m}}^{I,\mathfrak{r},Q},\widetilde{\mathfrak{m}}_Q,\chi},$$

obtained *before* taking coinvariants under the centre. The component groups of our Shimura varieties are indexed by a finite abelian (ray) class group $C = F^\times\backslash\mathbf{A}_F^\times/U$ for some U. The action of $\gamma \in A_F^\times$ on components is via the class $[\gamma]^2$, and the action of the central character on our cohomology groups is via $|\cdot|^2$ times a character of C. Let $C = C_p \oplus C^p$, where C_p is the p-Sylow subgroup of C. There are always natural isomorphisms of $\mathcal{O}[\Delta_Q]$ modules

$$\widetilde{M}^{\chi,I,\mathfrak{r}} \simeq M^{\chi,I,\mathfrak{r}} \otimes_{\mathcal{O}} \mathcal{O}[C_p],$$

$$\widetilde{M}^{\chi,I,\mathfrak{r},Q} \simeq M^{\chi,I,\mathfrak{r},Q} \otimes_{\mathcal{O}} \mathcal{O}[C_p],$$

with Δ_Q acting trivially on the second factor. The reason for such an isomorphism is that, after localization at a maximal ideal $\mathfrak{m}$ of the Hecke algebra, the elements of the centre which act through an element of order prime to p are already determined, because they are fixed modulo $\mathfrak{m}$ and the polynomial $T^m - 1$ is separable modulo p if $(m,p) = 1$. On the other hand, if we consider only the connected components corresponding to the subgroup C^p, the entire space is canonically isomorphic to $|C_p|$ copies of this space, and moreover, the action of $C/C^p = C_p$ on these components is transitive and fixed point free (this crucially uses that $p \neq 2$). Hence working with the $|\cdot|^2$ part of the cohomology is simply equivalent to working with the components indexed by C^p instead of C, and the passage between the cohomology (or complexes) for either of these two spaces (even before localization) is simply to tensor with $\mathcal{O}[C_p]$.

Part (1) follows immediately from these considerations, because $M^{\bullet,I,\chi}_{K^p_1(Q)}$ is perfect of amplitude $[0,0]$. By Lemma 2.10.2, to prove part (2), it is enough to prove that the corresponding perfect complex (of amplitude $[0,1]$) has Euler characteristic 0 after localization at $\widetilde{\mathfrak{m}}^{I,\mathfrak{r}}$. We can check this modulo any prime ideal of Λ_I, so the result follows from Theorem 6.6.5 and Corollary 3.10.5. □

7.9. **Taylor–Wiles systems: local-global compatibility.** We write $\mathbf{T}^{\chi,I,\mathfrak{r}}$ for the Λ_I-subalgebra of $\mathrm{End}_{\Lambda_I}(M^{\chi,I,\mathfrak{r}})$ generated by the image of $\widetilde{\mathbf{T}}^I$. Similarly, we write $\mathbf{T}^{\chi,I,\mathfrak{r},Q}$ for the Λ_I-subalgebra of $\mathrm{End}_{\Lambda_I}(M^{\chi,I,\mathfrak{r},Q})$ generated by the image of $\widetilde{\mathbf{T}}^{I,Q}$. We remind the reader that none of these objects depend on the choice of $\mathfrak{r}_v$ for $v \in I$ (but they do depend on the choice of $\mathfrak{r}_v$ for $v \notin I$). If $v \in I$, then by Hensel's lemma and our assumption that $\overline{\alpha}_v \neq \overline{\beta}_v$, we can write

$$X^2 - U_{\mathrm{Kli}(v),1}X + U_{v,2} = (X - \tilde{\alpha}_v)(X - \tilde{\beta}_v)$$

where $\tilde{\alpha}_v, \tilde{\beta}_v \in \mathbf{T}^{\chi,I,\mathfrak{r},Q}$ are respectively lifts of $\overline{\alpha}_v, \overline{\beta}_v$.

If $I \subset I'$, then there is a natural surjective map $\Lambda_I \to \Lambda_{I'}$, corresponding to the closed immersion $\mathrm{Spec}\,\Lambda_{I'} \to \mathrm{Spec}\,\Lambda_I$ given by $\theta_{v,1} = \theta_{v,2}$ for all $v \in I'$. Then we have the following key doubling statement:

Proposition 7.9.1 (Doubling). *For each choice of $\mathfrak{s}$, and each $I \subset I'$, there are natural surjections*

$$M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'} \to M^{\chi,I',\mathfrak{s}}$$

and

$$M^{\chi,I,\mathfrak{s},Q} \otimes_{\Lambda_I} \Lambda_{I'} \to M^{\chi,I',\mathfrak{s},Q}$$

which commute with all the Hecke operators away from $I' \setminus I$. Furthermore, if $v \in I' \setminus I$, then these surjections are equivariant with respect to $U_{v,0}$ and $U_{v,2}$, and intertwine the actions of $U_{v,1}$ on the source and $\widetilde{\mathfrak{s}}_v$ on the target.

Proof of Proposition 7.9.1. We give the proof for $M^{\chi,I,\mathfrak{s}}$, as the argument for $M^{\chi,I,\mathfrak{s},Q}$ is identical. By induction, it suffices to consider the case that $I' = I \cup \{v\}$ for some $v \notin I$. We have a map of complexes

$$M^{\bullet,I'}_{K_p} \to M^{\bullet,I}_{K_p},$$

induced by the restriction map coming from the inclusion

$$\mathfrak{X}^{G_1,I}_{K,\mathrm{Kli}(p^\infty)} \hookrightarrow \mathfrak{X}^{G_1,I'}_{K,\mathrm{Kli}(p^\infty)},$$

together with the natural map $\Lambda_I \to \Lambda_{I'}$. This induces a map

$$M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'} \xrightarrow{\tau} M^{\chi,I',\mathfrak{s}},$$

and the map that we are seeking is the map $\tau \circ U_{v,1} - \widetilde{\mathfrak{s}}'_v \circ \tau$. It is clear that this satisfies all of the claimed properties except possibly for the surjectivity and the claimed intertwining of $U_{v,1}$ and $\widetilde{\mathfrak{s}}_v$.

To see the intertwining, it is convenient to introduce the module $M^{\chi,I',\mathfrak{s},=_v 2}$, whose definition is

$$M^{\chi,I',\mathfrak{s},=_v 2} = \mathrm{RHom}^0_{\Lambda_{I'}}(M^{\bullet,I}_{K^p} \otimes_{\Lambda_I} \Lambda_{I'}, \Lambda_{I'})_{\widetilde{\mathfrak{m}}^{I',\mathfrak{s}},\chi,|\cdot|^2};$$

that is, it is defined in the same way as $M^{\chi,I,\mathfrak{s}}$, but we are now over the weight space $\Lambda_{I'}$, rather than Λ_I, and we localize with respect to the Hecke operator $(U_{\mathrm{Kli}(v),1} - (\alpha_v + \beta_v))$, rather than $(U_{\mathrm{Iw}(v),1} - \mathfrak{s}_v)$. By Lemma 4.5.17, on $M^{\chi,I',\mathfrak{s},=_v 2}$ we have the identity

$$U_{v,1}(U_{\mathrm{Kli}(v),1} - U_{v,1}) = U_{v,2},$$

or equivalently (writing $\{\alpha_v, \beta_v\} = \{\mathfrak{s}_v, \mathfrak{s}'_v\}$) the identity

$$(U_{v,1} - \widetilde{\mathfrak{s}}'_v)(U_{v,1} - \widetilde{\mathfrak{s}}_v) = 0. \tag{7.9.2}$$

We need to show that $U_{v,1} = \widetilde{\mathfrak{s}}'_v$ on $M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'}$. Now, noting that $M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'}$ is a subspace of $M^{\chi,I',\mathfrak{s},=_v 2}$ (because it is obtained from it by localizing with respect to $(U_{\mathrm{Iw}(v),1} - \mathfrak{s}_v)$, and because (7.9.2) holds on $M^{\chi,I',\mathfrak{s},=_v 2}$), we see that (7.9.2) also holds on $M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'}$; since $U_{\mathrm{Iw}(v),1}$ acts via $\mathfrak{s}_v$ modulo the maximal ideal of $\mathbf{T}^{\chi,I,\mathfrak{s}}$, it follows from Hensel's lemma that $U_{v,1} = \mathfrak{s}_v$ on $M^{\chi,I,\mathfrak{s}} \otimes_{\Lambda_I} \Lambda_{I'}$, as required.

It only remains to check the surjectivity. By Nakamaya's lemma, it is enough to check surjectivity modulo $\mathfrak{m}_{\Lambda_{I'}}$, or equivalently (by Lemma 7.8.5) the injectivity of

the map

$$\begin{array}{c} e(U^{I'})H^0(X^{I',G_1}_{K^pK_p(I'),1},\omega^2(-D))_{\tilde{\mathfrak{m}}^{I',\mathfrak{s}},\chi,|\cdot|^2} \\ \Big\downarrow{\scriptstyle (U_{v,1}-\mathfrak{s}'_v)} \\ e(U^I)H^0(X^{I,G_1}_{K_p(I)K^p,1},\omega^2(-D))_{\tilde{\mathfrak{m}}^{I,\mathfrak{s}},\chi,|\cdot|^2} \end{array} \tag{7.9.3}$$

on the special fibre. This follows from Theorem 5.8.6, as in Remark 5.8.7. □

Recall from §7.3 that if $v \in I$, we defined a character $\theta_v : I_{F_v} \to \Lambda_v^\times$, and if $v \notin I$ we defined a pair of characters $\theta_{v,1}, \theta_{v,2} : I_{F_v} \to \Lambda_v^\times$. We extend all of these characters to G_{F_v} by sending $\mathrm{Art}_{F_v}(p) \mapsto 1$. In the following theorem, we allow the Taylor–Wiles datum $(Q, (\overline{\alpha}_{v,1}, \dots, \overline{\alpha}_{v,4})_{v\in Q})$ to be empty.

Theorem 7.9.4. *There is a unique continuous representation*

$$\rho^{\chi,I,\mathfrak{s},Q} : G_F \to \mathrm{GSp}_4(\mathbf{T}^{\chi,I,\mathfrak{s},Q})$$

which is a deformation of $\overline{\rho}$ of type $\mathcal{S}^{I,\mathfrak{s}}_{\chi,Q}$ such that the induced homomorphism $R_{\mathcal{S}^{I,\mathfrak{s}}_{\chi,Q}} \to \mathbf{T}^{\chi,I,\mathfrak{s},Q}$ is a homomorphism of $\Lambda_I[\Delta_Q]$-algebras, and moreover such that

(1) If $v \notin S_p \cup R \cup \{v_0\} \cup Q$, then $\det(X - \rho^{\chi,I,\mathfrak{s},Q}(\mathrm{Frob}_v)) = Q_v(X)$.

(2) If $v \in I$, then

$$\rho^{\chi,I,\mathfrak{s},Q}|_{G_{F_v}} \simeq \begin{pmatrix} \lambda_{\tilde{\alpha}_v}\theta_v & 0 & * & * \\ 0 & \lambda_{\tilde{\beta}_v}\theta_v & * & * \\ 0 & 0 & \lambda^{-1}_{\tilde{\beta}_v}\theta_v^{-1}\varepsilon^{-1} & 0 \\ 0 & 0 & 0 & \lambda^{-1}_{\tilde{\alpha}_v}\theta_v^{-1}\varepsilon^{-1} \end{pmatrix}.$$

(3) If $v \in I^c$, then

$$\rho^{\chi,I,\mathfrak{s},Q}|_{G_{F_v}} \simeq \begin{pmatrix} \lambda_{U_{v,1}}\theta_{v,1} & * & * & * \\ 0 & \lambda_{U_{v,2}/U_{v,1}}\theta_{v,2} & * & * \\ 0 & 0 & \lambda^{-1}_{U_{v,2}/U_{v,1}}\theta_{v,2}^{-1}\varepsilon^{-1} & * \\ 0 & 0 & 0 & \lambda^{-1}_{U_{v,1}}\theta_{v,1}^{-1}\varepsilon^{-1} \end{pmatrix}.$$

Proof. First we treat the case $I = \emptyset$. By Proposition 7.8.11, $M^{\chi,\emptyset,\mathfrak{s},Q}$ is a finite free $\Lambda_\emptyset$-module, so there is an injection of $\widetilde{\mathbf{T}}^{\emptyset,Q}$-modules

$$M^{\chi,\emptyset,\mathfrak{s},Q} \to \prod_\kappa M^{\chi,\emptyset,\mathfrak{s},Q} \otimes_{\Lambda_\emptyset,\kappa} E$$

where the product is over all weights $\kappa = (k_v, l_v)_{v|\infty}$ with $k_v \geq l_v \geq 4$, $k_v \equiv l_v \equiv 2$ or $p+1 \pmod{2(p-1)}$. (Note that these points are scheme-theoretically dense in $\mathrm{Spec}\,\Lambda_\emptyset$.)

From the definition of $M^{\chi,\emptyset,\mathfrak{s},Q}$, Lemma 7.8.5, and Theorem 6.6.5, we have

$$(M^{\chi,\emptyset,\mathfrak{s},Q} \otimes_{\Lambda,\kappa} E)^\vee = e(\emptyset)H^0(X^{G_1}_{K_1^p(Q)K_p(\emptyset)},\omega^\kappa)^{\{T(k(v))=\chi_v\}_{v\in R},|\cdot|^2}_{\tilde{\mathfrak{m}}^{I,\mathfrak{s},Q}},$$

and by Theorem 3.10.1, we have

$$H^0(X^{G_1}_{K_1^p(Q)K_p(\emptyset)},\omega^\kappa) \otimes \overline{E} \simeq \bigoplus_\pi (\pi_f^{K_p^1(Q)K_p(\emptyset)}) \otimes \overline{E},$$

where in the sum, π runs over all the cuspidal automorphic representations of weight (k_v, l_v), with π_v holomorphic for each $v|\infty$, and π_f is the finite part of π.

Next we observe that for such a π, if the $\widetilde{\mathbf{T}}^{\emptyset,Q}$-module

$$e(\emptyset)(\pi_f^{K_1^p(Q)K_p(\emptyset)} \otimes \overline{E})_{\widetilde{\mathfrak{m}}^{I,\tilde{s},Q}}^{\{T(k(v))=\chi_v\}_{v\in R},|\cdot|^2}$$

is nonzero, then π has central character $|\cdot|^2$ and by Proposition 2.4.26, π is ordinary, and moreover $\widetilde{\mathbf{T}}^{\emptyset,Q}$ acts on it through a character $\Theta_\pi : \widetilde{\mathbf{T}}^{\emptyset,Q} \to \overline{E}$, and the ordinary Hecke parameters are $(\Theta_\pi(U_{v,1}), \Theta_\pi(U_{v,2}/U_{v,1}))$.

We now argue as in the proof of [CHT08, Prop. 3.4.4]. By Theorem 2.7.2, Proposition 2.4.13, Proposition 2.4.28, Proposition 2.4.30, and Remark 2.4.31, there is a Galois representation $\rho_{\pi,p} : G_F \to \mathrm{GSp}_4(\overline{E})$ such that

- If $v \notin S_p \cup R \cup \{v_0\} \cup Q$, then $\rho_{\pi,p}|_{G_{F_v}}$ is unramified and $\det(X - \rho_{\pi,I}(\mathrm{Frob}_v)) = \Theta_\pi(Q_v(X))$.
- If $v \in S_p$, then

$$\rho_{\pi,p}|_{G_{F_v}} \simeq \begin{pmatrix} \lambda_{\Theta_\pi(U_{v,1})}\theta_{v,1} & * & * & * \\ 0 & \lambda_{\Theta_\pi(U_{v,2}/U_{v,1})}\theta_{v,2} & * & * \\ 0 & 0 & \lambda^{-1}_{\Theta_\pi(U_{v,2}/U_{v,1})}\theta^{-1}_{v,2}\varepsilon^{-1} & * \\ 0 & 0 & 0 & \lambda^{-1}_{\Theta_\pi(U_{v,1})}\theta^{-1}_{v,1}\varepsilon^{-1} \end{pmatrix}$$

- If $v \in R$, then for all $\sigma \in I_{F_v}$, $\det(X - \rho(\sigma))$ is equal to

$$(X - \chi_{v,1}(\mathrm{Art}_{F_v}^{-1}(\sigma)))(X - \chi_{v,1}(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}(X - \chi_{v,2}(\sigma))(X - \chi_{v,2}(\mathrm{Art}_{F_v}^{-1}(\sigma))^{-1}).$$

- If $v \in Q$, then

$$\rho|_{G_{F_v}} \simeq \gamma_{v,1} \oplus \gamma_{v,2} \oplus \varepsilon^{-1}\gamma_{v,2}^{-1} \oplus \varepsilon^{-1}\gamma_{v,1}^{-1}$$

  for characters $\gamma_{v,i} : G_{F_v} \to \overline{E}^\times$ satisfying $\overline{\gamma}_i = \lambda_{\overline{\alpha}_{v,i}}$. Furthermore $T(F_v)$ acts on $(\pi^{\mathrm{Iw}_1(v)})_{\mathfrak{m}_{\overline{\alpha}_1,\overline{\alpha}_2}}$ via the characters $\gamma_{v,i} \circ \mathrm{Art}_{F_v}$.

After conjugation, we may assume that $\rho_{\pi,p}$ is valued in $\mathcal{O}_{E_\pi}$ for some finite extension E_π/E, and since $\overline{\rho}_{\pi,p} \cong \overline{\rho}$, we may assume after further conjugation that $\overline{\rho}_{\pi,p} = \overline{\rho}$. Let A be the subring of $k \oplus \bigoplus_\pi \mathcal{O}_{E_\pi}$ consisting of those elements $(a, (a_\pi)_\pi) \in k \oplus \bigoplus_\pi \mathcal{O}_{E_\pi}$ such that for all π the reduction of a_π modulo the maximal ideal of $\mathcal{O}_{E_\pi}$ is equal to a (where the direct sum is over the infinitely many π corresponding to the infinitely many κ). Then A is a local Λ-algebra with residue field k (with the Λ-algebra structure coming from that on $\mathcal{O}_{E_\pi}$ given by κ). Set

$$\rho_A := \overline{\rho} \oplus \bigoplus_\pi \rho_{\pi,p} : G_F \to \mathrm{GSp}_4(A).$$

There is a natural injection $\mathbf{T}^{\chi,\emptyset,\tilde{s},Q} \to A$ (this map is injective because $\mathbf{T}^{\chi,\emptyset,\tilde{s},Q}$ is reduced, by a standard argument using Proposition 2.4.26). We can choose (for example, by ordering the κ) a decreasing sequence of ideals I_n of A with $\cap_n I_n = (0)$ such that each A/I_n is an object of CNL_Λ, and it follows from [GG12, Lem. 7.1.1] that for each n the representation $\rho_A \otimes_A A/I_n$ is $\ker(\mathrm{GSp}_4(A/I_n) \to \mathrm{GSp}_4(k))$-conjugate to a representation

$$\rho_n^{\chi,\emptyset,\tilde{s},Q} : G_F \to \mathrm{GSp}_4(\mathbf{T}^{\chi,\emptyset,\tilde{s},Q}/(I_n \cap \mathbf{T}^{\chi,\emptyset,\tilde{s},Q})).$$

After possibly conjugating again, we can assume that $\rho_{n+1}^{\chi,\emptyset,\tilde{s},Q} \pmod{I_n} = \rho_n^{\chi,\emptyset,\tilde{s},Q}$, and we set $\rho^{\chi,\emptyset,\tilde{s},Q} := \varprojlim_n \rho_n^{\chi,\emptyset,\tilde{s},Q}$. By construction this satisfies the required

properties at places $v \notin S_p \cup Q$ (in particular, at the places $v \in R$, the deformation is of the required type by the definition of R_v^χ).

It remains to verify the claimed properties of $\rho^{\chi,\emptyset,\bar{\mathfrak{r}},Q}|_{G_{F_v}}$ for $v \in S_p \cup Q$. Suppose that $v \in S_p$. We claim firstly that it is enough to show that there are elements $\nu_{v,1}, \nu_{v,2} \in (\mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q})^\times$ such that

$$\rho^{\chi,\emptyset,\bar{\mathfrak{r}},Q}|_{G_{F_v}} \cong \begin{pmatrix} \lambda_{\nu_{v,1}}\theta_{v,1} & * & * & * \\ 0 & \lambda_{\nu_{v,2}}\theta_{v,2} & * & * \\ 0 & 0 & \lambda_{\nu_{v,2}}^{-1}\theta_{v,2}^{-1}\varepsilon^{-1} & * \\ 0 & 0 & 0 & \lambda_{\nu_{v,1}}^{-1}\theta_{v,1}^{-1}\varepsilon^{-1} \end{pmatrix} \tag{7.9.5}$$

Indeed, if this holds, then the equalities $\nu_{v,1} = U_{v,1}$ and $\nu_{v,2} = U_{v,2}/U_{v,1}$ can be checked after composing with the injection $\mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q} \hookrightarrow A$, where they follow from local-global compatibility for the $\rho_{\pi,p}|_{G_{F_v}}$. Now, (7.9.5) is equivalent to asking that the homomorphism $R_v^\square \to \mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q}$ corresponding to $\rho^{\chi,\emptyset,\bar{\mathfrak{r}},Q}|_{G_{F_v}}$ factors through the quotient $R_v^{B,\bar{\mathfrak{r}}_v}$, and this can again be checked after composing with the injection $\mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q} \hookrightarrow A$, as required.

Suppose now that $v \in Q$, so that we need to check that the morphism $R_{\mathcal{S}_{\chi,Q}^{I,\bar{\mathfrak{r}}}} \to \mathbf{T}^{\chi,I,\bar{\mathfrak{r}},Q}$ is Δ_v-equivariant. By Lemma 7.4.4, there are unique characters $\gamma_{v,1}, \gamma_{v,2} : G_{F_v} \to (\mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q})^\times$ lifting $\lambda_{\overline{\alpha}_{v,1}}$, $\lambda_{\overline{\alpha}_{v,2}}$ respectively such that $\rho^{\chi,\emptyset,\bar{\mathfrak{r}},Q}|_{G_{F_v}} \cong \gamma_{v,1} \oplus \gamma_{v,2} \oplus \gamma_{v,2}\varepsilon^{-1} \oplus \gamma_{v,1}\varepsilon^{-1}$. We claim that the action of $T(F_v)$ on $M^{\chi,I,\bar{\mathfrak{r}},Q}$ is given by $\gamma_{v,1} \circ \mathrm{Art}_{F_v}$, $\gamma_{v,2} \circ \mathrm{Art}_{F_v}$; this can be checked after composing with the injection $\mathbf{T}^{\chi,\emptyset,\bar{\mathfrak{r}},Q} \hookrightarrow A$, so it follows from the analogous result for $\rho_{\pi,p}|_{G_{F_v}}$ recalled above. Restricting this claim to $T(\mathcal{O}_{F_v})$ gives the result.

We are done in the case that $I = \emptyset$. We now prove the result for general I by induction on $\#I$. Accordingly, assume that the result holds for some $I \neq S_p$, choose $w \in I^c$, and set $I' = I \cup \{w\}$. By Proposition 7.9.1 we have a natural surjection of $\Lambda_{I'}$-algebras

$$\widetilde{\mathbf{T}}^{\chi,I,\bar{\mathfrak{r}},Q} \otimes_{\Lambda_I} \Lambda_{I'} \twoheadrightarrow \widetilde{\mathbf{T}}^{\chi,I',\bar{\mathfrak{r}},Q},$$

and we let $\rho^{\chi,I',\bar{\mathfrak{r}},Q}$ be the pushforward of $\rho^{\chi,I,\bar{\mathfrak{r}},Q}$. It follows from the result for I that we need only check that property (2) holds for $v = w$. However, we could equally well have performed the same construction with $\bar{\mathfrak{r}}_w$ replaced with $\bar{\mathfrak{r}}'_w$ (the two candidates for $\rho^{\chi,I',\bar{\mathfrak{r}},Q}$ are conjugate by property (1), the Cebotarev density theorem, and [GG12, Lem. 7.1.1]), so from (3) and the equivariance properties for Hecke operators at w in Proposition 7.9.1, we see that $\rho^{\chi,I',\bar{\mathfrak{r}},Q}|_{G_{F_w}}$ admits both $\lambda_{\tilde{\alpha}_w}\theta_w$ and $\lambda_{\tilde{\beta}_w}\theta_w$ as subcharacters. Since $\overline{\alpha}_w \neq \overline{\beta}_w$, the result follows. □

As a corollary, we have the following result about Galois representations associated to automorphic representations of parallel weight 2. As ever, some of the hypotheses in this result could be relaxed (in particular, the assumption that $\overline{\rho}_{\pi,p}$ is vast and tidy can presumably easily be relaxed to irreducibility), but in the interests of brevity we have contented ourselves with this result, as it is sufficient for our purposes.

Corollary 7.9.6. *Let π be a cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 with central character $|\cdot|^2$. Fix a prime $p > 2$, and assume that π is ordinary. Then there is a continuous semisimple representation $\rho_{\pi,p} : G_F \to \mathrm{GL}_4(\overline{\mathbf{Q}}_p)$ such that*

(1) For each finite place $v \nmid p$, at which π_v is unramified, $\rho_{\pi,p}|_{G_{F_v}}$ is unramified and

$$\det(X - \rho_{\pi,p}(\mathrm{Frob}_v)) = Q_v(X).$$

Suppose further that $\overline{\rho}_{\pi,p}$ is vast and tidy, and that for each $v|p$, the ordinary Hecke parameters α_v, β_v of π_v satisfy $\overline{\alpha}_v \neq \overline{\beta}_v$. Then $\rho_{\pi,p}$ can be conjugated to be valued in $\mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$, and

(2) $\nu \circ \rho_{\pi,p} = \varepsilon^{-1}$.

(3) For each finite place $v \nmid p$, we have

$$\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}.$$

(4) For each place $v|p$, then

$$\rho_{\pi,p}|_{G_{F_v}} \cong \begin{pmatrix} \lambda_{\alpha_v} & 0 & * & * \\ 0 & \lambda_{\beta_v} & * & * \\ 0 & 0 & \lambda_{\beta_v}^{-1}\varepsilon^{-1} & 0 \\ 0 & 0 & 0 & \lambda_{\alpha_v}^{-1}\varepsilon^{-1} \end{pmatrix}.$$

Proof. This could be proved by repeating the arguments of [Mok14, §4], using Theorem 7.9.4 instead of the results of [MT15]. For brevity, we instead explain how to deduce the result from [Mok14, Thm. 4.14] and Theorem 7.9.4.

Firstly, if π is not of general type in the sense of [Art04], then the existence of a (unique) semisimple *reducible* representation $\rho_{\pi,p}$ satisfying (1) is an easy consequence of standard results on Galois representations for GL_1 and GL_2 (see the proof of Lemma 2.9.1), and parts (2)-(4) are then vacuous.

Accordingly, for the remainder of the proof we assume that π is of general type, in which case the existence of a representation $\rho_{\pi,p}$ satisfying (1) and (3) follows from [Mok14, Thm. 4.14], except that this representation is only given to be valued in $\mathrm{GL}_4(\overline{\mathbf{Q}}_p)$ rather than $\mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$.

Choose a solvable extension of totally real fields F'/F, linearly disjoint from $\overline{F}^{\ker \overline{\rho}_{\pi,p}}$ over F, with the properties that p splits completely in F', and that there is an automorphic representation Π of $\mathrm{GSp}_4(\mathbf{A}_{F'})$ of parallel weight 2 and central character $|\cdot|^2$, which is a base change of π (that is, for each finite place w of F', lying over a place v of F, we have $\mathrm{rec}_{\mathrm{GT},p}(\Pi) = \mathrm{rec}_{\mathrm{GT},p}(\pi)|_{W_{F'_w}}$), which is holomorphic at all infinite places, and which satisfies $\Pi_w^{\mathrm{Iw}(w)} \neq 0$ for all finite places w of F' (the existence of such an F' and Π follows from [Mok14, Prop. 4.13]).

We claim that if $\rho_{\Pi,p}$ admits a symplectic pairing with multiplier ε^{-1}, then so does $\rho_{\pi,p}$. Indeed, since $\rho_{\Pi,p} = \rho_{\pi,p}|_{G_{F'}}$ is irreducible, it admits at most one perfect pairing with multiplier ε^{-1}; while by (1), $\rho_{\pi,p}$ admits a perfect pairing with multiplier ε^{-1}, which must therefore also be symplectic. In addition (4) holds for $\rho_{\Pi,p}$ if and only if it holds for $\rho_{\pi,p}$. Replacing F by F' and π by Π, we can and do assume that $\pi_v^{\mathrm{Iw}(v)} \neq 0$ for all finite places v of F.

Taking $\overline{\rho} := \overline{\rho}_{\pi,p}$, we see that Hypothesis 7.8.1 holds, so the required properties of $\rho_{\pi,p}$ follow immediately from Theorem 7.9.4, taking $I = S_p$, $\chi = 1$ and $Q = \emptyset$. (Note that as in the proof of Theorem 7.9.4, it follows from Theorem 3.10.1 that π contributes to $M^{1,S_p,\hat{\imath},\emptyset}$.) □

We now turn to the final lemmas that we need to prove in order to construct our Taylor–Wiles systems.

Lemma 7.9.7. *Let $\Lambda \in \mathrm{CNL}_{\mathcal{O}}$, and let $f^\bullet : \mathcal{C}^\bullet \to \mathcal{D}^\bullet$ be a morphism of bounded complexes of $\mathfrak{m}_\Lambda$-adically complete and separated flat Λ-modules. Supposed that the induced morphism $\mathcal{C}^\bullet \otimes^{\mathbf{L}}_\Lambda \Lambda/\mathfrak{m}_\lambda \to \mathcal{D}^\bullet \otimes^{\mathbf{L}}_\Lambda \Lambda/\mathfrak{m}_\lambda$ is a quasi-isomorphism. Then $f^\bullet$ is a quasi-isomorphism.*

Proof. See [Pil20, Prop. 2.2]. □

Proposition 7.9.8. *The natural map $M^{\chi,I,\tilde{\tau},Q} \to M^{\chi,I,\tilde{\tau}}$ induces an isomorphism $(M^{\chi,I,\tilde{\tau},Q})_{\Delta_Q} \to M^{\chi,I,\tilde{\tau}}$.*

Proof. We follow the proof of [KT17, Lem. 6.25]. We claim that we have natural isomorphisms

$$(M^{\chi,I,\tilde{\tau},Q})_{\Delta_Q} \xrightarrow{\sim} M^{\chi,I,\tilde{\tau},Q}_{K_0^p(Q)} \tag{7.9.9}$$

and

$$M^{\chi,I,\tilde{\tau},Q}_{K_0^p(Q)} \xrightarrow{\sim} M^{\chi,I,\tilde{\tau}} \tag{7.9.10}$$

whose composite is the claimed isomorphism. We begin with (7.9.9). It suffices to show that we have a natural isomorphism in the derived category

$$(M^{\bullet,I}_{K_1^p(Q)})^{\prod_{v\in Q} T(k(v))} \xrightarrow{\sim} M^{\bullet,I}_{K_0^p(Q)}.$$

As in the proof of Proposition 7.8.11, the complex on the left (before taking invariants) is a perfect complex of $\Lambda_I[\prod_{v\in Q} T(k(v))]$-modules. But now the result is immediate from Proposition 7.8.10, as the map $X_{K_pK_1^p(Q),\Sigma} \to X_{K_pK_0^p(Q),\Sigma}$ is finite étale with group $\prod_{v\in Q} T(k(v))$.

We now turn to proving (7.9.10). Again, we mostly work on the level of complexes. We begin by considering the composite

$$(M^{\bullet,I}_{K_0^p(Q)})_{\tilde{\mathfrak{m}}^{\mathrm{an},Q},\tilde{\mathfrak{m}}_Q} \to (M^{\bullet,I}_{K_0^p(Q)})_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}} \to (M^{\bullet,I}_{K^p})_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}}.$$

By Lemma 7.9.7, these maps induce quasi-isomorphism of complexes if the following maps are isomorphisms

$$H^*(M^{\bullet,I}_{K_0^p(Q)} \otimes k)_{\tilde{\mathfrak{m}}^{\mathrm{an},Q},\tilde{\mathfrak{m}}_Q} \to H^*(M^{\bullet,I}_{K_0^p(Q)} \otimes k)_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}} \to H^*(M^{\bullet,I}_{K^p} \otimes k)_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}}.$$

This follows formally from Lemmas 2.4.36 and 2.4.37, applied at each place in Q, because, for $K = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ and $K' = \mathrm{Iw}(v)$, we have the identities of Hecke operators

$$[K1K'][K'1K] = [K : K']$$
$$[K'1K][K1K'] = e_K = e_{\mathrm{GSp}_4(\mathcal{O}_{F_v})},$$

and we note that $[K1K']$ is the trace from level K' to $\mathrm{GSp}_4(\mathcal{O}_{F_v})$ and $[K'1K]$ is the inclusion from level K to level K' (recall that since $p > 2$, $[K : K'] = [\mathrm{GSp}_4(\mathcal{O}_{F_v}) : \mathrm{Iw}(v)]$ is not divisible by p).

Finally, consider the natural map

$$(M^{\bullet,I}_{K^p})_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}} \to (M^{\bullet,I}_{K^p})_{\tilde{\mathfrak{m}}^{\mathrm{an}}}.$$

For our purposes, it suffices to prove that this map becomes an isomorphism after applying $\mathrm{RHom}^0(-,\Lambda)$. Since this map is a localisation, it suffices to check that it is an isomorphism modulo the maximal ideal of Λ; so by Lemma 7.8.5 (1), it is in turn enough to prove that

$$H^0(M^{\bullet,I}_{K^p} \otimes k)_{\tilde{\mathfrak{m}}^{\mathrm{an},Q}} \to H^0(M^{\bullet,I}_{K^p} \otimes k)_{\tilde{\mathfrak{m}}^{\mathrm{an}}}$$

is an isomorphism, or in other words, that $\widetilde{\mathfrak{m}}^{\mathrm{an}}$ is the unique maximal ideal $\mathfrak{n}$ of $\widetilde{\mathbf{T}}$ lying over $\widetilde{\mathfrak{m}}^{\mathrm{an},Q}$ and in the support of $H^0(M^{\bullet,I}_{K^p} \otimes k)$. Equivalently, we need to show that the Hecke eigenvalues away from the primes in Q (which are prime to the level) determine the Hecke eigenvalues at Q. This follows from the fact that the Hecke eigenvalues at primes of good reduction and residue characteristic different from p are determined by the Galois representation (exactly as in the proof of Theorem 7.9.4, this local-global compatibility statement for H^0 is a consequence of the corresponding local-global compatibility statement for the Galois representations in Theorem 2.7.2). But the Galois representation itself is determined from $\widetilde{\mathfrak{m}}^{\mathrm{an},Q}$ by the Cebotarev density theorem. Hence $\mathfrak{n} = \widetilde{\mathfrak{m}}^{\mathrm{an}}$, as required. □

7.10. **An abstract patching criterion.** We have the following slight variant on [CG18, Prop. 2.3, Prop. 6.6] (although our formulation is also informed by [KT17, Prop. 3.1]); we leave the details of the proof as an exercise for the interested reader.

Proposition 7.10.1. *Let l_0 be equal to either 0 or 1, let $\Lambda \in \mathrm{CNL}_{\mathcal{O}}$, let $S_\infty := \Lambda[[x_1, \dots, x_q]]$ for some $q \geq 1$, and set $\mathbf{a} := \ker(S_\infty \to \Lambda)$. Let $S_\infty \supset I_1 \supset I_2 \supset \dots$ be a decreasing sequence of open ideals of S_∞ with $\cap_N I_N = 0$. For each $N \geq 1$ we set $S_N = S_\infty / I_N$.*

Suppose that we are given the following data.

- *Objects R^1_∞, R^χ_∞ of CNL_Λ.*
- *Objects R^1, R^χ of CNL_Λ, an R^1-module M^1, and an R^χ-module M^χ, each of which is finite as a Λ-module. Furthermore if $l_0 = 0$, then they are both free as Λ-modules, and if $l_0 = 1$, then they are balanced Λ-modules.*
- *For each integer $N \geq 1$, finite S_N-modules M^1_N, M^χ_N, which are free if $l_0 = 0$ and balanced if $l_0 = 1$, together with isomorphisms of S_N-modules $M^1_N/\mathbf{a} \xrightarrow{\sim} M^1 \otimes_{S_\infty} S_N$, $M^\chi_N/\mathbf{a} \xrightarrow{\sim} M^\chi \otimes_{S_\infty} S_N$ (where the action of S_∞ on M^1, M^χ is via the augmentation $S_\infty \to \Lambda$).*
- *For each $N \geq 1$, objects R^1_N, R^χ_N of CNL_{S_N}, and maps of S_N-algebras $R^1_N \to R^1/I_N$, $R^\chi_N \to R^\chi/I_N$ and $R^1_N \to \mathrm{End}_{S_N}(M^1_N)$, $R^\chi_N \to \mathrm{End}_{S_N}(M^\chi_N)$, such that the two following diagrams commute.*

$$\begin{tikzcd} R^1_N \arrow[r] \arrow[d] & \mathrm{End}_{S_N}(M^1_N) \arrow[d] \\ R^1/I_N \arrow[r] & \mathrm{End}_{\Lambda/I_N}(M^1 \otimes_{S_\infty} S_N) \end{tikzcd}$$

$$\begin{tikzcd} R^\chi_N \arrow[r] \arrow[d] & \mathrm{End}_{S_N}(M^\chi_N) \arrow[d] \\ R^\chi/I_N \arrow[r] & \mathrm{End}_{\Lambda/I_N}(M^\chi \otimes_{S_\infty} S_N) \end{tikzcd}$$

- *For each $N \geq 1$, surjections of Λ-algebras $R^1_\infty \twoheadrightarrow R^1_N$, $R^\chi_\infty \twoheadrightarrow R^\chi_N$.*

We suppose also that we are given the following compatibilities between the data indexed by 1 and the data indexed by χ.

- *isomorphisms of Λ/λ-algebras $R^1_\infty/\lambda \cong R^\chi_\infty/\lambda$, $R^1/\lambda \cong R^\chi/\lambda$, and $R^1_N/\lambda \cong R^\chi_N/\lambda$, compatible with the surjections $R^1_\infty \twoheadrightarrow R^1_N$ and $R^\chi_\infty \twoheadrightarrow R^\chi_N$.*
- *An isomorphism of $R^1/\lambda \cong R^\chi/\lambda$-modules $M^1/\lambda \cong M^\chi/\lambda$.*

- *For each $N \geq 1$, isomorphisms of S_N/λ-modules $M_N^1/\lambda \cong M_N^\chi/\lambda$, compatible with all actions, and such that the following diagram commutes, where we write J_N for the kernel of the composite $\Lambda \to S_\infty \to S_\infty/I_N$.*

$$\begin{tikzcd} M_N^1/(\lambda, \mathbf{a}) \arrow[r] \arrow[d] & M_N^\chi/(\lambda, \mathbf{a}) \arrow[d] \\ M^1/(\lambda, J_N) \arrow[r] & M^\chi/(\lambda, J_N) \end{tikzcd}$$

Then we can find the following data.

- *Homomorphisms of Λ-algebras $S_\infty \to R_\infty^1$, $S_\infty \to R_\infty^\chi$.*
- *Finite S_∞-modules M_∞^1, M_∞^χ, which are free if $l_0 = 0$ and balanced if $l_0 = 1$, together with isomorphisms $M_\infty^1 \otimes_{S_\infty} \Lambda \xrightarrow{\sim} M^1$, and $M_\infty^\chi \otimes_{S_\infty} \Lambda \xrightarrow{\sim} M^\chi$.*
- *Commutative diagrams of S_∞-algebras*

$$\begin{tikzcd} R_\infty^1 \arrow[r] \arrow[d] & \mathrm{End}_{S_\infty}(M_\infty^1) \arrow[d, "-\otimes_{S_\infty}\Lambda"] \\ R^1 \arrow[r] & \mathrm{End}_\Lambda(M^1) \end{tikzcd}$$

$$\begin{tikzcd} R_\infty^\chi \arrow[r] \arrow[d] & \mathrm{End}_{S_\infty}(M_\infty^\chi) \arrow[d, "-\otimes_{S_\infty}\Lambda"] \\ R^\chi \arrow[r] & \mathrm{End}_\Lambda(M^\chi) \end{tikzcd}$$

- *An isomorphism $M_\infty^1/\lambda \xrightarrow{\sim} M_\infty^\chi/\lambda$, compatible with the actions of $R_\infty^1/\lambda \xrightarrow{\sim} R_\infty^\chi/\lambda$, such that the following diagram commutes.*

$$\begin{tikzcd} M_\infty^1/(\lambda, \mathbf{a}) \arrow[r] \arrow[d] & M_\infty^\chi/(\lambda, \mathbf{a}) \arrow[d] \\ M^1/\lambda \arrow[r] & M^\chi/\lambda \end{tikzcd}$$

7.11. **The patching construction.** We now apply Proposition 7.10.1 to our spaces of p-adic automorphic forms. We continue to assume that Hypothesis 7.8.1 holds.

Enlarging E if necessary, we can and do assume that E contains a primitive pth root of unity, and a primitive 9th root of unity if $p = 3$. By Hypothesis 7.8.1 (4a), for each $v \in R$ we can and do choose a pair of non-trivial characters $\chi_v = (\chi_{v,1}, \chi_{v,2})$, with $\chi_{v,i} : \mathcal{O}_{F_v}^\times \to \mathcal{O}^\times$ which are trivial modulo λ, and such that $\chi_{v,1} \neq \chi_{v,2}^{\pm 1}$. We will now apply the constructions of the previous sections, simultaneously using both this choice of χ, and also the choice $\chi = 1$. In the former case we will label our objects as we did before, and in the latter we will replace χ by 1.

Let

$$q = h^1(F_S/F, \mathrm{ad}^0\,\overline{\rho}(1)), \qquad g = 2q - 4[F : \mathbf{Q}] + \#S - 1,$$

and set $\Delta_\infty = \mathbf{Z}_p^{2q}$. Let $S_\infty = \mathcal{T}[[\Delta_\infty]]$, where $\mathcal{T}$ is as in §7.1. Viewing S_∞ as an augmented Λ-algebra, we let $\mathbf{a}$ denote the augmentation ideal.

For each $N \geq 1$, we fix a choice of Taylor–Wiles datum $(Q_N, (\overline{\alpha}_{v,1}, \ldots, \overline{\alpha}_{v,4})_{v \in Q_N})$ as in Corollary 7.6.3. For $N = 0$, we set $Q_0 = \emptyset$. For each $N \geq 1$, we let $\Delta_N = \Delta_{Q_N} = \prod_{v \in Q_N} k(v)^\times(p)^2$ and fix a surjection $\Delta_\infty \twoheadrightarrow \Delta_N$. The kernel of

this surjection is contained in $(p^N\mathbf{Z}_p)^{2q}$, since each $v \in Q_N$ satisfies $q_v \equiv 1 \bmod p^N$. We let Δ_0 be the trivial group, viewed as a quotient of Δ_∞. We write $S_N = \mathcal{T}[\Delta_N]$.

For each $N \geq 0$, we set $R_N^{1,I,\mathfrak{r}} = R_{\mathcal{S}_{1,Q_N}^{I,\mathfrak{r}}}$ and $R_N^{\chi,I,\mathfrak{r}} = R_{\mathcal{S}_{\chi,Q_N}^{I,\mathfrak{r}}}$. Note that $R_0^{1,I,\mathfrak{r}} = R_{\mathcal{S}_1}^{I,\mathfrak{r}}$ and $R_0^{\chi,I,\mathfrak{r}} = R_{\mathcal{S}_\chi}^{I,\mathfrak{r}}$. Let $R^{1,I,\mathfrak{r},\mathrm{loc}} = R_{\mathcal{S}_1^{I,\mathfrak{r}}}^{S,\mathrm{loc}}$ and $R^{\chi,I,\mathfrak{r},\mathrm{loc}} = R_{\mathcal{S}_\chi^{I,\mathfrak{r}}}^{S,\mathrm{loc}}$ denote the corresponding completed tensor product of local deformation rings, as in §7.2. By definition we have

$$R^{1,I,\mathfrak{r},\mathrm{loc}} = (\widehat{\otimes}_{v\in I} R_v^P)\widehat{\otimes}(\widehat{\otimes}_{v\in I^c} R_v^{B,\bar{\mathfrak{r}}_v})\widehat{\otimes}(\widehat{\otimes}_{v\in R} R_v^1)\widehat{\otimes} R_{v_0}^{\square},$$

$$R^{\chi,I,\mathfrak{r},\mathrm{loc}} = (\widehat{\otimes}_{v\in I} R_v^P)\widehat{\otimes}(\widehat{\otimes}_{v\in I^c} R_v^{B,\bar{\mathfrak{r}}_v})\widehat{\otimes}(\widehat{\otimes}_{v\in R} R_v^{\chi_v})\widehat{\otimes} R_{v_0}^{\square},$$

with all completed tensor products being taken over $\mathcal{O}$.

For any $N \geq 1$, we have $R_{\mathcal{S}_{1,Q_N}^{I,\mathfrak{r}}}^{S,\mathrm{loc}} = R^{1,I,\mathfrak{r},\mathrm{loc}}$ and $R_{\mathcal{S}_{\chi,Q_N}^{I,\mathfrak{r}}}^{S,\mathrm{loc}} = R^{\chi,I,\mathfrak{r},\mathrm{loc}}$. There are canonical isomorphisms $R^{1,I,\mathfrak{r},\mathrm{loc}}/(\lambda) \cong R^{\chi,I,\mathfrak{r},\mathrm{loc}}/(\lambda)$ and $R_N^{1,I,\mathfrak{r}}/(\lambda) \cong R_N^{\chi,I,\mathfrak{r}}/(\lambda)$ for all $N \geq 0$. For each $N \geq 1$, $R_N^{1,I,\mathfrak{r}}$ and $R_N^{\chi,I,\mathfrak{r}}$ are canonically $\Lambda[\Delta_N]$-algebras and there are canonical isomorphisms $R_N^{1,I,\mathfrak{r}} \otimes_{\Lambda[\Delta_N]} \Lambda \cong R_0^{1,I,\mathfrak{r}}$ and $R_N^{\chi,I,\mathfrak{r}} \otimes_{\Lambda[\Delta_N]} \Lambda \cong R_0^{\chi,I,\mathfrak{r}}$, which are compatible with the isomorphisms modulo λ.

Fix representatives $\rho_{\mathcal{S}_\chi^{I,\mathfrak{r}}}$, $\rho_{\mathcal{S}_1^{I,\mathfrak{r}}}$ of the universal deformations which are identified modulo λ (via the identifications $R_{\mathcal{S}_\chi^{I,\mathfrak{r}}}/(\lambda) \cong R_{\mathcal{S}_1^{I,\mathfrak{r}}}/(\lambda)$). By Lemma 7.1.6, these give rise to an $R^{1,I,\mathfrak{r},\mathrm{loc}}$-algebra structure on $R_N^{1,I,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$ and an $R^{\chi,I,\mathfrak{r},\mathrm{loc}}$-algebra structure on $R_N^{\chi,I,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$; the canonical isomorphism $R^{1,I,\mathfrak{r},\mathrm{loc}}/(\lambda) \cong R^{\chi,I,\mathfrak{r},\mathrm{loc}}/(\lambda)$ is compatible with these algebra structures and with the canonical isomorphisms $R_N^{1,I,\mathfrak{r}}/(\lambda) \cong R_N^{\chi,I,\mathfrak{r}}/(\lambda)$. We let $R_\infty^{1,I,\mathfrak{r}}$ and $R_\infty^{\chi,I,\mathfrak{r}}$ be formal power series rings in g variables over $R^{1,I,\mathfrak{r},\mathrm{loc}}$ and $R^{\chi,I,\mathfrak{r},\mathrm{loc}}$, respectively. By Proposition 7.2.1 and Corollary 7.6.3, we can choose local Λ-algebra surjections $R_\infty^{1,I,\mathfrak{r}} \to R_N^{1,I,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$ and $R_\infty^{\chi,I,\mathfrak{r}} \to R_N^{\chi,I,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$ for every $N \geq 0$. We can and do assume that these are compatible with our fixed identifications modulo λ, and with the natural isomorphisms $R_N^{1,I,\mathfrak{r}} \otimes_{\Lambda[\Delta_N]} \Lambda \cong R_0^{1,I,\mathfrak{r}}$ and $R_N^{\chi,I,\mathfrak{r}} \otimes_{\Lambda[\Delta_N]} \Lambda \cong R_0^{\chi,I,\mathfrak{r}}$.

Fix a subset $I \subset S_p$ of cardinality $\#I \leq 1$, and a choice of $\mathfrak{r}$. We now apply Proposition 7.10.1, taking (in the notation established in §7.7):

- Λ to be Λ_I.
- S_∞, S_N to be as above.
- $R_\infty^1 := R_\infty^{1,I,\mathfrak{r}}$, $R_\infty^\chi := R_\infty^{\chi,I,\mathfrak{r}}$, $R^1 := R_0^{1,I,\mathfrak{r}}$, $R^\chi := R_0^{\chi,I,\mathfrak{r}}$, $R_N^1 := R_N^{1,\infty,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$, $R_N^\chi := R_N^{\chi,\infty,\mathfrak{r}}\widehat{\otimes}_\Lambda\mathcal{T}$.
- $M^1 := M^{1,I,\mathfrak{r}}$, $M^\chi := M^{\chi,I,\mathfrak{r}}$, $M_N^1 := M^{1,I,\mathfrak{r},Q_N}\widehat{\otimes}_\Lambda\mathcal{T}$, $M_N^\chi := M^{\chi,I,\mathfrak{r},Q_N}\widehat{\otimes}_\Lambda\mathcal{T}$.

By Theorem 7.9.4, Proposition 7.9.8 and Proposition 7.8.11, this data satisfies the assumptions of Proposition 7.10.1. Consequently, we have:

- Λ_I-algebra homomorphisms $S_\infty \to R_\infty^{1,I,\mathfrak{r}}$ and $S_\infty \to R_\infty^{\chi,I,\mathfrak{r}}$.
- Finite S_∞-modules $M_\infty^{1,I,\mathfrak{r}}$, $M_\infty^{\chi,I,\mathfrak{r}}$ which are free if $\#I = 0$ and balanced if $\#I = 1$, together with isomorphisms $M_\infty^{1,I,\mathfrak{r}}/\mathbf{a} \cong M^{1,I,\mathfrak{r}}$, $M^{\chi,I,\mathfrak{r}}/\mathbf{a} \cong M^{\chi,I,\mathfrak{r}}$.
- Morphisms of S_∞-algebras $R_\infty^{1,I,\mathfrak{r}} \to \mathrm{End}_{S_\infty}(M^{1,I,\mathfrak{r}})$, $R_\infty^{\chi,I,\mathfrak{r}} \to \mathrm{End}_{S_\infty}(M^{\chi,I,\mathfrak{r}})$, which are compatible with the actions of R^1, R^χ on $M^{1,I,\mathfrak{r}}$, $M^{\chi,I,\mathfrak{r}}$ respectively.
- Isomorphisms

$$M_\infty^{1,I,\mathfrak{r}}/\lambda M_\infty^{1,I,\mathfrak{r}} \simeq M_\infty^{\chi,I,\mathfrak{r}}/\lambda M_\infty^{\chi,I,\mathfrak{r}}, \quad M^{1,I,\mathfrak{r}}/\lambda M^{1,I,\mathfrak{r}} \simeq M^{\chi,I,\mathfrak{r}}/\lambda M^{\chi,I,\mathfrak{r}}$$

compatible with the actions of $R_\infty^{1,I,\mathfrak{s}}/(\lambda) \simeq R_\infty^{\chi,I,\mathfrak{s}}/(\lambda)$ and $R^{1,I,\mathfrak{s}}/(\lambda) \simeq R^{\chi,I,\mathfrak{s}}/(\lambda)$ and the above isomorphisms.

We now briefly pause to introduce some notation that will be in force throughout the rest of §7. We will need to work with $\mathcal{O}$-flat modules M over complete local Noetherian $\mathcal{O}$-algebras R which are not necessarily $\mathcal{O}$-flat, but for which we have good control of $R[1/p]$. There are various ways that we could do this, but we have found it convenient to reduce to the $\mathcal{O}$-flat case in the following way. For a Noetherian complete local $\mathcal{O}$-algebra R we denote by R' the maximal $\mathcal{O}$-flat quotient of R (i.e. the image of R in $R[1/p]$, or equivalently the quotient of R by its ideal of p-power torsion). Note that if M is an R-module that is $\mathcal{O}$-flat then it is naturally an R'-module.

Returning to the situation at hand, by definition, S_∞ is formally smooth over Λ_I of relative dimension $2q+11\#S-1$, and Λ_I is formally smooth over $\mathcal{O}$. By Propositions 7.3.4, 7.4.7, 7.4.8, and 7.4.2, and [BLGHT11, Lem. 3.3], $(R_\infty^{1,I,\mathfrak{s}})'$ and $(R_\infty^{\chi,I,\mathfrak{s}})'$ are equidimensional of relative dimension $g+10\#S+4[F:\mathbf{Q}]-\#I$ over Λ_I. By the definition of g, we conclude that

$$\dim(R_\infty^{1,I,\mathfrak{s}})' = \dim(R_\infty^{\chi,I,\mathfrak{s}})' = \dim S_\infty - \#I. \tag{7.11.1}$$

Proposition 7.11.2. *$M_\infty^{1,I,\mathfrak{s}}$ is a maximal Cohen–Macaulay $(R_\infty^{1,I,\mathfrak{s}})'$-module, and $M_\infty^{\chi,I,\mathfrak{s}}$ is a maximal Cohen–Macaulay $(R_\infty^{\chi,I,\mathfrak{s}})'$-module.*

Proof. These statements have identical proofs, so we give the argument for the first of them. From (7.11.1), we see that the support of $M_\infty^{1,I,\mathfrak{s}}$ in $\operatorname{Spec} S_\infty$ has codimension at least $\#I$. By [CG18, Lem. 6.2] (applied to a resolution $S_\infty^r \to S_\infty^r$ of $M_\infty^{1,I,\mathfrak{s}}$ if $\#I=1$ — such a resolution exists, by Lemma 2.10.2 — and to $M_\infty^{1,I,\mathfrak{s}}$ itself if $\#I=0$), we see that the codimension is precisely $\#I$, and that M_∞^1 has depth $\dim S_\infty - \#I = \dim(R_\infty^{1,I,\mathfrak{s}})'$ over S_∞. It follows that the depth of $M_\infty^{1,I,\mathfrak{s}}$ over $(R_\infty^{1,I,\mathfrak{s}})'$ is at least $\dim(R_\infty^{1,I,\mathfrak{s}})'$, so that $M_\infty^{1,I,\mathfrak{s}}$ is maximal Cohen–Macaulay over $(R_\infty^{1,I,\mathfrak{s}})'$, as required. □

7.12. **Cycles and modules over products of local deformation rings.** In preparation for our study of the dimensions of certain spaces of p-adic modular forms in the next section, we formalize some arguments which are at the heart of our version of the "Ihara avoidance" argument of [Tay08]. Following [EG14], we use the language of cycles on the special fibres of (completed tensor products of) local deformation rings; our perspective is also informed by [Sho18].

We recall some notation for cycles and multiplicities from [EG14, §2]. In particular, if R is an equidimensional Noetherian local ring of dimension d then by a cycle (or a d-cycle) on $\operatorname{Spec} R$ we mean simply a formal $\mathbf{Z}$-linear combination of the generic points of $\operatorname{Spec} R$. We denote the group of cycles on R by $\mathcal{Z}^d(R)$ (or just $\mathcal{Z}(R)$, with the understanding that we will only consider top-dimensional cycles). If M is a finite R-module then the cycle of M is defined by

$$Z(M,R) = \sum_\eta \operatorname{len}_{R_\eta}(M_\eta)\cdot \eta$$

where the sum is over the generic points η of $\operatorname{Spec} R$ and $\operatorname{len}_{R_\eta}(M_\eta)$ denotes the length of M_η as a R_η-module.

If R is an equidimensional, flat, Noetherian $\mathcal{O}$-algebra of dimension $d+1$ and η is a generic point of $\operatorname{Spec} R$ then we write R^η for the quotient of R by the minimal

prime corresponding to η, and we let $\overline{\eta} = Z(R^\eta/(\lambda), R/(\lambda))$. Then [EG14, Prop. 2.2.13] states that if M is a finite R-module which is $\mathcal{O}$-flat, then

$$Z(M/\lambda M, R/(\lambda)) = \sum_\eta \mathrm{len}_{R_\eta}(M_\eta) \cdot \overline{\eta}$$

where the sum is over the generic points η of R.

Next we recall several facts about completed tensor products. As in §7.11, if $R \in \mathrm{CNL}_{\mathcal{O}}$, we let R' denote the maximal p-torsion free quotient of R. Let $R_1, R_2 \in \mathrm{CNL}_{\mathcal{O}}$. First we note that the natural map $R_1 \widehat{\otimes} R_2 \to R_1' \widehat{\otimes} R_2'$ induces an isomorphism $(R_1 \widehat{\otimes} R_2)' \simeq R_1' \widehat{\otimes} R_2'$. (Indeed this follows from the fact that the kernel is p-power torsion and that $R_1' \widehat{\otimes} R_2'$ is $\mathcal{O}$-flat, see [Tho15, Lem. 1.3].)

Now suppose that R_1 and R_2 are $\mathcal{O}$-flat and equidimensional of dimensions d_1+1 and d_2+1 respectively, and further assume that all the irreducible components of $\operatorname{Spec} R_i$ and $\operatorname{Spec} R_i/(\lambda)$ for $i = 1, 2$ are geometrically irreducible (for instance by enlarging $\mathcal{O}$ if necessary). Write $R = R_1 \widehat{\otimes} R_2$; then R is $\mathcal{O}$-flat and equidimensional of dimension $d_1 + d_2 + 1$. (This, and the other facts recalled in this paragraph, can be read off from [Tho15, Lem 1.4].) Moreover if η_i is a generic point of $\operatorname{Spec} R_i$ for $i = 1, 2$ then the kernel of the natural map

$$R \to R_1^{\eta_1} \widehat{\otimes} R_2^{\eta_2}$$

is a minimal prime of R which corresponds to a generic point of $\operatorname{Spec} R$ which we denote by $\eta = (\eta_1, \eta_2)$, and the generic points of $\operatorname{Spec} R$ are precisely the (η_1, η_2) as η_i ranges over the generic points of $\operatorname{Spec} R_i$ for $i = 1, 2$. Similarly if $\mathfrak{p}_i \subset R_i/(\lambda)$ is a minimal prime for $i = 1, 2$ then

$$(\mathfrak{p}_1, \mathfrak{p}_2) = \ker\big(R/(\lambda) \to R_1/(\lambda, \mathfrak{p}_1) \widehat{\otimes} R_2/(\lambda, \mathfrak{p}_2)\big)$$

is a minimal prime of $R/(\lambda)$, and every minimal prime of $R/(\lambda)$ has this form. It follows that there is an isomorphism

$$\begin{aligned} \mathcal{Z}^{d_1}(R_1/(\lambda)) \otimes \mathcal{Z}^{d_2}(R_2/(\lambda)) &\to \mathcal{Z}^{d_1+d_2}(R/(\lambda)), \\ \eta_1 \otimes \eta_2 &\mapsto (\eta_1, \eta_2). \end{aligned}$$

According to [EG14, Lem. 2.2.14], if M_i is a finite R_i-module for $i = 1, 2$, so that we may form the R-module $M = M_1 \widehat{\otimes}_{\mathcal{O}} M_2$, then under the above isomorphism we have

$$Z(M_1/\lambda M_1, R_1/(\lambda)) \otimes Z(M_2/\lambda M_2, R_2/(\lambda)) = Z(M/\lambda M, R/(\lambda)). \tag{7.12.1}$$

In particular for a generic point $\eta = (\eta_1, \eta_2)$ of R we have an isomorphism $R^\eta \simeq R^{\eta_1} \widehat{\otimes} R^{\eta_2}$ of R-modules and hence, in the notation introduced above, under this isomorphism we have $\overline{\eta} = \overline{\eta}_1 \otimes \overline{\eta}_2$.

We wish to apply this discussion to the rings

$$R^1 = \widehat{\otimes}_{v \in R} R_v^1, \quad R^\chi = \widehat{\otimes}_{v \in R} R_v^\chi$$

as well as to $\widetilde{R}^1 = R^1 \widehat{\otimes} \widetilde{R}$ and $\widetilde{R}^\chi = R^\chi \widehat{\otimes} \widetilde{R}$, for some auxiliary $\widetilde{R} \in \mathrm{CNL}_{\mathcal{O}}$ with the property that $\widetilde{R}'$ is irreducible. (In applications $\widetilde{R}^1$ and $\widetilde{R}^\chi$ will be $R_\infty^{1,I,\mathfrak{s}}$ and $R_\infty^{\chi,I,\mathfrak{s}}$ for some choice of I and $\mathfrak{s}$; so $\widetilde{R}$ is formally smooth over a completed tensor product of the deformation rings considered in Proposition 7.3.4, and $\widetilde{R}'$ is indeed irreducible.)

We recall that for each $v \in R$ we have $R_v^1/(\lambda) = R_v^\chi/(\lambda)$. Passing to p-torsion free quotients, it is not the case that $(R_v^1)'/(\lambda)$ is identified with $(R_v^\chi)'/(\lambda)$, but

Propositions 7.4.7 and 7.4.8 imply that at least the underlying topological spaces of $\mathrm{Spec}(R_v^1)'/(\lambda)$ and $\mathrm{Spec}(R_v^\chi)'/(\lambda)$ coincide with that of $\mathrm{Spec}\, R_v^1/(\lambda) = \mathrm{Spec}\, R_v^\chi/(\lambda)$, and so in particular $\mathcal{Z}((R_v^1)'/(\lambda)) = \mathcal{Z}((R_v^\chi)'/(\lambda))$. Passing to products we obtain identifications $\mathcal{Z}((R^1)'/(\lambda)) = \mathcal{Z}((R^\chi)'/(\lambda))$ and $\mathcal{Z}((\widetilde{R}^1)'/(\lambda)) = \mathcal{Z}((\widetilde{R}^\chi)'/(\lambda))$.

Lemma 7.12.2. *Let M^1 be a finite $\mathcal{O}$-flat $\widetilde{R}^1$-module, and let M^χ be a finite $\mathcal{O}$-flat $\widetilde{R}^\chi$-module, such that $M^1/\lambda M^1 \simeq M^\chi/\lambda M^\chi$ as $\widetilde{R}^1/(\lambda) = \widetilde{R}^\chi/(\lambda)$-modules. Then*

$$Z(M^1/\lambda M^1, (\widetilde{R}^1)'/(\lambda)) = Z(M^\chi/\lambda M^\chi, (\widetilde{R}^\chi)'/(\lambda))$$

under the identification of $\mathcal{Z}((\widetilde{R}^1)'/(\lambda))$ with $\mathcal{Z}((\widetilde{R}^\chi)'/(\lambda))$ from above.

Proof. As we explained above we have two quotients $(\widetilde{R}^1)'/(\lambda)$ and $(\widetilde{R}^\chi)'/(\lambda)$ of $\widetilde{R}^1/(\lambda) = \widetilde{R}^\chi/(\lambda)$ whose spectra have the same underlying topological space. Each generic point of this space corresponds to minimal primes $\mathfrak{p}^1$ and $\mathfrak{p}^\chi$ of $(\widetilde{R}^1)'/(\lambda)$ and $(\widetilde{R}^\chi)'/(\lambda)$ as well as to a (not necessarily minimal) prime $\mathfrak{p}$ of $\widetilde{R}^1/(\lambda) = \widetilde{R}^\chi/(\lambda)$ which is the preimage of both $\mathfrak{p}^1$ and $\mathfrak{p}^\chi$. Then we claim that we have equalities

$$\mathrm{len}_{((\widetilde{R}^1)'/(\lambda))_{\mathfrak{p}^1}}((M^1/\lambda M^1)_{\mathfrak{p}^1}) = \mathrm{len}_{(\widetilde{R}^1/(\lambda))_{\mathfrak{p}}}((M^1/\lambda M^1)_{\mathfrak{p}}) = \mathrm{len}_{((\widetilde{R}^\chi)'/(\lambda))_{\mathfrak{p}^\chi}}((M^\chi/\lambda M^\chi)_{\mathfrak{p}^\chi})$$

which exactly gives the statement of the lemma. Both equalities follow from the fact that if $A \to B$ is a surjective map of rings and M is a finite length B-module then $\mathrm{len}_A(M) = \mathrm{len}_B(M)$. □

For the next lemma we need to introduce some more notation. From a tuple $\eta = (\eta_v)_{v\in R}$ of generic points η_v of $\mathrm{Spec}(R_v^1)'$ for $v \in R$ we obtain a generic point η of $\mathrm{Spec}(R^1)'$ (resp. a generic point also denoted η of $\mathrm{Spec}(\widetilde{R}^1)'$) and moreover these are all of the generic points of $\mathrm{Spec}(R^1)'$ (resp. of $\mathrm{Spec}(\widetilde{R}^1)'$). By Proposition 7.4.7, if $\eta_{v,1}$ and $\eta_{v,2}$ are two distinct generic points of $\mathrm{Spec}(R_v^1)'$ for some $v \in R$, then the cycles $\overline{\eta}_{v,1}$ and $\overline{\eta}_{v,2}$ have disjoint support.

It follows from this and (7.12.1) that if η_1 and η_2 are two distinct generic points of $\mathrm{Spec}(R^1)'$ (resp. of $\mathrm{Spec}(\widetilde{R}^1)'$) then the supports of $\overline{\eta}_1$ and $\overline{\eta}_2$ are disjoint. Finally recall that by Proposition 7.4.8, for each $v \in R$, $\mathrm{Spec}(R_v^\chi)'$ is irreducible. Passing to products, $\mathrm{Spec}(R^\chi)'$ and $\mathrm{Spec}(\widetilde{R}^\chi)'$ are irreducible as well. We denote the unique generic point of either by η^χ.

As already indicated, in the statement and proof of the following lemma, we freely identify the generic points of $\mathrm{Spec}(R^1)'$ and $\mathrm{Spec}(\widetilde{R}^1)'$ (and we also identify the generic points of $\mathrm{Spec}(R^1)'/(\lambda)$ and $\mathrm{Spec}(\widetilde{R}^1)'/(\lambda)$).

Lemma 7.12.3. *Suppose there exists a finite, $\mathcal{O}$-flat $\widetilde{R}^1$-module M^1, and a finite, $\mathcal{O}$-flat $\widetilde{R}^\chi$-module M^χ, along with an isomorphism $M^1/\lambda M^1 \simeq M^\chi/\lambda M^\chi$ of $\widetilde{R}^1/(\lambda) = \widetilde{R}^\chi/(\lambda)$-modules. Suppose furthermore that M^1 is supported on at least one generic point η of $\mathrm{Spec}(\widetilde{R}^1)'$. Then there exist unique positive integers d'_η labelled by generic points $\eta = (\eta_v)_{v\in R}$ of $\mathrm{Spec}(R^1)'$ such that*

(1) As elements of $\mathcal{Z}((R^1)'/(\lambda)) = \mathcal{Z}((R^\chi)'/(\lambda))$ we have

$$\overline{\eta^\chi} = \sum_\eta d'_\eta \overline{\eta} \tag{7.12.4}$$

where the sum is over the generic points η of $\mathrm{Spec}(R^1)'$.

(2) For each generic point η of $\mathrm{Spec}(\widetilde{R}^1)'$ *we have*

$$\mathrm{len}_{(\widetilde{R}^1)'_\eta}(M^1_\eta) = d'_\eta \,\mathrm{len}_{(\widetilde{R}^\chi)'_{\eta^\chi}}(M^\chi_\eta). \tag{7.12.5}$$

In particular M^1 is supported on every generic point of $\mathrm{Spec}(\widetilde{R}^1)'$.

Proof. As explained above, the cycles $\overline{\eta}$ as η ranges over the generic points of $\mathrm{Spec}(R^1)'$ have disjoint support. Thus the formula (7.12.4) uniquely determines the integers d'_η. Moreover, as the cycle $\overline{\eta^\chi}$ is supported on every generic point of $\mathrm{Spec}(R^\chi)'/(\lambda)$, (7.12.4) also implies that the integers d_η must be positive, if they exist.

Now using [EG14, Prop. 2.2.13] as recalled above, we have

$$Z(M^1/\lambda M^1, (\widetilde{R}^1)'/(\lambda)) = \sum_\eta \mathrm{len}_{(\widetilde{R}^1)'_\eta} M^1_\eta \cdot \overline{\eta}$$

and

$$Z(M^\chi/\lambda M^\chi, (\widetilde{R}^\chi)'/(\lambda)) = \mathrm{len}_{(\widetilde{R}^\chi)'_{\eta^\chi}} M^\chi_{\eta^\chi} \cdot \overline{\eta^\chi}$$

and moreover these two cycles coincide by Lemma 7.12.2.

Our hypothesis that M^1 is supported on some generic point η of $(\widetilde{R}^1)'$ implies that $\mathrm{len}_{(\widetilde{R}^1)'_\eta} M^1_\eta > 0$. Hence by the above equality of cycles, $\mathrm{len}_{(\widetilde{R}^\chi)'_{\eta^\chi}} M^\chi_{\eta^\chi} > 0$. Because the cycles $\overline{\eta}$ have disjoint support, we must have that

$$d'_\eta = \frac{\mathrm{len}_{(\widetilde{R}^1)'_\eta}(M^1_\eta)}{\mathrm{len}_{(\widetilde{R}^\chi)'_{\eta^\chi}}(M^\chi_{\eta^\chi})}$$

is an integer for each generic point η of $\mathrm{Spec}(R^1)'$, and for this choice of d'_η, the formulas (7.12.4) and (7.12.5) hold. □

Remark 7.12.6. We note that the "multiplicities" d'_η in Lemma 7.12.3 are independent of the modules M^1 and M^χ and even of the auxiliary ring $\widetilde{R}$. In the next section they will be given a local representation-theoretic interpretation (see Proposition 7.13.5 and Remark 7.13.12).

Remark 7.12.7. In §7.13 we will use Lemma 7.12.3 to compute the dimensions of spaces of p-adic modular forms at Iwahori level. The idea of comparing patched modules over $R^{1,I,\mathfrak{s}}_\infty$ and $R^{\chi,I,\mathfrak{s}}_\infty$ goes back to [Tay08]; the key point is that $R^{\chi,I,\mathfrak{s}}_\infty[1/p]$ is a domain, which guarantees that the support of an appropriate patched module is all of $\mathrm{Spec}\, R^{\chi,I,\mathfrak{s}}_\infty$, and the isomorphism $R^{1,I,\mathfrak{s}}_\infty/(\lambda) = R^{\chi,I,\mathfrak{s}}_\infty/(\lambda)$ which allows us to transfer this information to $R^{1,I,\mathfrak{s}}_\infty$.

7.13. **Multiplicities of patched spaces of p-adic automorphic forms.** We now make use of our patching constructions to determine the multiplicities of systems of eigenvalues corresponding to ρ in spaces of p-adic automorphic forms with $\#I \leq 1$.

We begin by introducing some notation and assumptions. We suppose that we have fixed a representation $\rho : G_F \to \mathrm{GSp}_4(\mathcal{O})$, which satisfies the following properties. (While this list of properties may appear to be too restrictive to be useful, we will later use base change to reduce to this situation.)

Hypothesis 7.13.1.

(1) F is a totally real field in which the prime $p \geq 3$ splits completely; we write S_p for the set of primes of F dividing p.

(2) $\nu \circ \rho = \varepsilon^{-1}$.
(3) For each finite place v of F, $\rho|_{G_{F_v}}$ is pure.
(4) For each $v \in S_p$, $\rho|_{G_{F_v}}$ is p-distinguished weight 2 ordinary, with unit eigenvalues $\alpha_v, \beta_v \in E$.
(5) There is a finite set R of primes of F not dividing p such that if $v \notin R \cup S_p$, then $\rho|_{G_{F_v}}$ is unramified, while if $v \in R$, then:
   - $q_v \equiv 1 \pmod p$, and if $p = 3$ then further $q_v \equiv 1 \pmod 9$.
   - $\overline{\rho}|_{G_{F_v}}$ is trivial.
   - $\rho|_{G_{F_v}}$ has only unipotent ramification.
(6) There exists $\pi = \otimes_v \pi_v$ an ordinary cuspidal automorphic representation for GSp_4/F of parallel weight 2 and central character $|\cdot|^2$, such that $\overline{\rho}_{\pi,p} = \overline{\rho}$, and such that:
   - For all $v \notin R \cup S_p$, π_v is unramified.
   - For all $v \in R \cup S_p$, $\pi_v^{\mathrm{Iw}(v)} \neq 0$.
   - For each finite place v of F, $\rho_{\pi,p}|_{G_{F_v}}$ is pure.
(7) The representation $\overline{\rho}$ is vast and tidy.

Remark 7.13.2. Note in particular that Hypothesis 7.13.1 implies that Hypothesis 7.8.1 holds for $\overline{\rho}$.

As in §7.7, it follows from Hypothesis 7.13.1, and in particular from the hypothesis that $\overline{\rho}(G_F)$ is tidy, that:

(8) There exists an absolutely unramified prime $v_0 \notin S_p \cup R$ with $q_{v_0} \not\equiv 1 \pmod p$ and residue characteristic greater than 5, such that $\rho|_{G_{F_{v_0}}}$ is unramified, and $\rho(\mathrm{Frob}_{v_0})$ has (not necessarily distinct) eigenvalues with the property that no ratio of these eigenvalues is congruent to q_{v_0} modulo λ.

Given a closed point $x \in \operatorname{Spec} R^{1,I,\mathfrak{r}}[1/p]$ or $x \in \operatorname{Spec} R^{\chi,I,\mathfrak{r}}[1/p]$, we will always assume that E is large enough to contain the residue field of x, so that in particular x parameterizes a Galois representation $\rho_x : G_F \to \mathrm{GSp}_4(\mathcal{O})$. We denote by $\mathfrak{p}_x$ the height one prime ideal which is the kernel of the corresponding homomorphism $R^{1,I,\mathfrak{r}} \to E$ or $R^{\chi,I,\mathfrak{r}} \to E$, and we also use the same symbol $\mathfrak{p}_x$ for the ideals obtained by pulling back under the homomorphisms $R^{1,I,\mathfrak{r},\mathrm{loc}} \to R^{1,I,\mathfrak{r}}_\infty \to R^{1,I,\mathfrak{r}}$ or under the homomorphisms $R^{\chi,I,\mathfrak{r},\mathrm{loc}} \to R^{\chi,I,\mathfrak{r}}_\infty \to R^{\chi,I,\mathfrak{r}}$. As in §7.4, we say that x (or ρ_x or $\mathfrak{p}_x$) is *smooth* if $R^{1,I,\mathfrak{r},\mathrm{loc}}_{\mathfrak{p}_x}$ (resp. $R^{\chi,I,\mathfrak{r},\mathrm{loc}}_{\mathfrak{p}_x}$) is a regular local ring (note that this is equivalent to their completions being regular).

For a Galois representation $\rho' : G_F \to \mathrm{GSp}_4(\mathcal{O})$ giving rise to a point x' on one of the deformation rings $R^{1,I,\mathfrak{r}}$ or $R^{\chi,I,\mathfrak{r}}$, we let $\tilde{\mathfrak{p}}^{\mathrm{an}}_{x'} \subset \widetilde{\mathbf{T}}$ be the corresponding prime ideal. Explicitly, this is the prime ideal generated by the coefficients of the polynomials $Q_v(X) - \det(X - \rho'(\mathrm{Frob}_v))$ for $v \notin R \cup S_p \cup \{v_0\}$. (As before, the "an" stands for "anaemic".)

For any choice of I and $\mathfrak{r}$, we let $\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_{x'} \subset \widetilde{\mathbf{T}}^I$ denote the prime ideal

(7.13.3)
$$\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_{x'} = (\tilde{\mathfrak{p}}^{\mathrm{an}}_{x'}, \{U_{v,0} - 1, U_{v,2} - \alpha'_v\beta'_v\}_{v \in S_p}, \{U_{\mathrm{Kli}(v),1} - \alpha'_v - \beta'_v\}_{v \in I}, \{U_{v,1} - \mathfrak{r}'_v\}_{v \in I^c}),$$

where, for $v \in S_p$, $\alpha'_v \equiv \overline{\alpha}_v \pmod \lambda$, $\beta'_v \equiv \overline{\beta}_v \pmod \lambda$ and $\mathfrak{r}'_v \in \{\alpha'_v, \beta'_v\}$ are determined by the local representations $\rho'|_{G_{F_v}}$ as in §7.3.

Definition 7.13.4. Let $K^{p,\mathrm{Iw}} = \prod_{v \nmid p,\infty} K_v$ and $K^{p,\mathrm{Iw}_1} = \prod_{v \nmid p,\infty} K'_v$, where

- $K_v = \mathrm{Iw}(v)$ and $K'_v = \mathrm{Iw}_1(v)$ for $v \in R$.

- $K_{v_0} = K'_{v_0} = \mathrm{Iw}_1(v_0)$.
- $K_v = K'_v = \mathrm{GSp}_4(\mathcal{O}_{F_v})$ for $v \notin R \cup \{v_0\}$.

We now define some spaces of p-adic modular forms. For any $I \subset S_p$, $\mathfrak{r}$, classical algebraic weight κ, and choice of $K^{p,\mathrm{Iw}}$ as in Definition 7.13.4, we let

$$S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}}K_p(I)} = (e(U^I)H^0(\mathfrak{X}^{I,G_1}_{K_p(I)K^p}, \omega^\kappa(-D))_{\tilde{\mathfrak{m}}^{I,\mathfrak{r}}}) \otimes_{\mathcal{O}} E[\{U_{v,0}-1\}_{v\in S_p}, \\ \{U_{v,0}-q_v^{-2}\}_{v\in R}, \{T_{v,0}-q_v^{-2}\}_{v\notin S_p\cup R\cup\{v_0\}}]$$

We also let

$$S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}_1}K_p(I),\chi} := (S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}_1}K_p(I)})^{\prod_{v\in R}\mathrm{Iw}(v)/\mathrm{Iw}_1(v)=\chi}$$

be the subspace with "nebentypus" corresponding to χ. By Lemma 7.8.5 (2), we have isomorphisms

$$(M^{1,I,\mathfrak{r}}/\mathfrak{p}_\kappa M^{1,I,\mathfrak{r}})[1/p] \simeq (S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}}K_p(I)})^\vee,$$
$$(M^{\chi,I,\mathfrak{r}}/\mathfrak{p}_\kappa M^{\chi,I,\mathfrak{r}})[1/p] \simeq (S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}_1}K_p(I),\chi})^\vee.$$

In particular, with this notation in place, for a Galois representation $\rho' : G_F \to \mathrm{GSp}_4(\mathcal{O})$ giving rise to a point on one of the deformation rings $R^{1,I,\mathfrak{r}}$ or $R^{\chi,I,\mathfrak{r}}$ and of weight κ (i.e. such that the composition $\Lambda_I \to R^{1,I,\mathfrak{r}} \to E$ or $\Lambda_I \to R^{\chi,I,\mathfrak{r}} \to E$ is κ) we have

$$(M^{1,I,\mathfrak{r}}_\infty/\mathfrak{p}_{x'}M^{1,I,\mathfrak{r}}_\infty)[1/p] \simeq (M^{1,I,\mathfrak{r}}/\mathfrak{p}_{x'}M^{1,I,\mathfrak{r}})[1/p] \simeq (S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_{x'}])^\vee,$$
$$(M^{\chi,I,\mathfrak{r}}_\infty/\mathfrak{p}_{x'}M^{\chi,I,\mathfrak{r}}_\infty)[1/p] \simeq (M^{1,I,\mathfrak{r}}/\mathfrak{p}_{x'}M^{1,I,\mathfrak{r}})[1/p] \simeq (S^{I,\mathfrak{r}}_{\kappa,K^{p,\mathrm{Iw}_1}K_p(I),\chi}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_{x'}])^\vee.$$

In order to state our results on the dimensions of eigenspaces of p-adic automorphic forms, we need to make a further study of the local deformation rings at places $v \in R$.

Proposition 7.13.5. *Let η_v be a generic point of* $\operatorname{Spec} R^1_v$ *for some $v \in R$. The set of $y \in (\operatorname{Spec} R^{1,\eta_v}_v)(\overline{E})$ such that the L-packet $L(\rho_y)$ contains a generic representation is nonempty, and the number*

$$d_{\eta_v} = \sum_{\pi\in L(\rho_y)} \dim \pi^{\mathrm{Iw}(v)}$$

is independent of such a y. More explicitly, the rank $n(\eta_v)$ of the monodromy operator N is generically constant on $\operatorname{Spec} R^{1,\eta_v}_v$*, and*

- *if $n(\eta_v) = 0$, then $d_\eta = 8$;*
- *if $n(\eta_v) = 1$, then $d_\eta = 4$;*
- *if $n(\eta_v) = 2$, then $d_\eta = 4$;*
- *if $n(\eta_v) = 3$, then $d_\eta = 1$.*

Proof. This can be read off from [RS07b, Tables A.7, A.15] (note that the rank of the monodromy operator is given in the column of [RS07b, Table A.15] headed "a"; note also that the unipotent L-packets which contain supercuspidal representations also contain generic non-supercuspidal representations, namely those of type Va and XIa, see [RS07c, §1], and [RS07b, Table A.1]). We see that:

- On the unramified components (those with $n(\eta_v) = 0$), the L-packets containing a generic representation are singletons $\{\pi\}$ of type I (unramified principal series), so $d_{\eta_v} = 8$.

- If $n(\eta_v) = 1$, the L-packets containing a generic representation are singletons $\{\pi\}$ of type IIa, so $d_{\eta_v} = 4$.
- If $n(\eta_v) = 2$, the L-packets containing a generic representation are either singletons $\{\pi\}$ of type IIIa, or pairs $\{\pi_a, \pi_b\}$ of respective types VIa and VIb, and in either case $d_{\eta_v} = 4$. (Note that the representations of type Va do not contribute, as they never correspond to residually trivial Galois representations.)
- Finally, if $n(\eta_v) = 3$, then the L-packets containing a generic representation are singletons of type IVa (unramified twists of Steinberg) and $d_{\eta_v} = 1$. □

We write $\eta = (\eta_v)_{v \in R}$ for a tuple of generic points η_v of $\operatorname{Spec} R_v^1$ for $v \in R$, which as explained in §7.12 gives rise to a generic point, also denoted η, of $R^{1,I,\mathfrak{s},\mathrm{loc}}$ or of $R_\infty^{1,I,\mathfrak{s}}$. We let

$$d_\eta = \prod_{v \in R} d_{\eta_v}$$

where d_{η_v} is as in Proposition 7.13.5. We also let $d_\rho = d_\eta$ for the generic point η of $R^{1,I,\mathfrak{s},\mathrm{loc}}$ that the local representations of ρ lie on (this point is unique, as the representations $\rho|_{G_{F_v}}$ are pure by assumption). Concretely, by Proposition 7.13.5, we have

$$d_\rho = 8^{|R_0|} 4^{|R_1|+|R_2|}$$

where for $i = 0, 1, 2, 3$, $R_i \subset R$ is the set of primes $v \in R$ for which $n(\rho|_{G_{F_v}}) = i$.

We can now state our main result about p-adic modularity at Iwahori level.

Theorem 7.13.6. *Assume Hypothesis 7.13.1 for $\rho = \rho_x$. For any I with $\#I \leq 1$, and any choice of $\mathfrak{s}$, we have*

$$\dim_E S^{I,\mathfrak{s}}_{\kappa_2, K^{p,\mathrm{Iw}} K_p(I)}[\tilde{\mathfrak{p}}_x^{I,\mathfrak{s}}] = 8 d_\rho.$$

Remark 7.13.7. The reason for the factor of $8 = |W|$ on the right hand side is that we are working at Iwahori level at the auxiliary place v_0, and not imposing any conditions on the Hecke operators at this place. It would be possible to impose such conditions and remove this factor, but we have found it more convenient not to do so (and it makes no difference for our main automorphy lifting theorems).

Before proving the theorem we recall a standard lemma, essentially due independently to Diamond and Fujiwara (see e.g. [Dia97]) which is the key to proving "multiplicity one" (or "multiplicity $8d_\rho$") results in characteristic 0 using the Taylor–Wiles method.

Lemma 7.13.8. *Let R be either $(R_\infty^{1,I,\mathfrak{s}})'$ or $(R_\infty^{\chi,I,\mathfrak{s}})'$ for some choice of I and $\mathfrak{s}$ and let M be a maximal Cohen–Macaulay R-module. Let $x \in \operatorname{Spec} R[1/p]$ be a smooth closed point with residue field E, and let $\mathfrak{p}_x \subset R$ be the corresponding prime ideal. Then $M_{\mathfrak{p}_x}$ is a free $R_{\mathfrak{p}_x}$-module, and hence if η is the unique generic point of R specializing to x, then*

$$\dim_{R_\eta} M_\eta = \dim_E (M/\mathfrak{p}_x M)[1/p].$$

Proof. The first statement follows from the fact that a maximal Cohen–Macaulay module over a regular local ring is free, and the second statement is an immediate consequence of this freeness. □

We also record the following proposition on "doubling":

Proposition 7.13.9. *Let $\rho = \rho_x$. For any choice of $I \subset S_p$, $w \in I^c$, K^p as in Definition 7.13.4, and $\mathfrak{r}$, there is an injection*

$$(U_{w,1} - \mathfrak{r}'_w) : S^{I\cup\{w\},\mathfrak{r}}_{\kappa_2,K^pK_p(I\cup\{w\})}[\tilde{\mathfrak{p}}_x^{I\cup\{w\},\mathfrak{r}}] \to S^{I,\mathfrak{r}}_{\kappa_2,K^pK_p(I)}[\tilde{\mathfrak{p}}_x^{I,\mathfrak{r}}].$$

Proof. This immediately reduces to the corresponding statement with $\mathcal{O}$-coefficients, and hence to the injectivity of the map (7.9.3) (with $w = v$), which we proved in the course of the proof of Proposition 7.9.1. □

We are now ready to prove Theorem 7.13.6.

Proof of Theorem 7.13.6. We first consider the case that $I = \emptyset$. As $M^{1,\emptyset,\mathfrak{r}}$ is a maximal Cohen–Macaulay $(R^{1,\emptyset,\mathfrak{r}})'$-module, it is supported on some irreducible component of $\mathrm{Spec}(R^{1,\emptyset,\mathfrak{r}})'$ and hence we may apply Lemma 7.12.3 to the $R^{1,\emptyset,\mathfrak{r}}_\infty$-module $M^{1,\emptyset,\mathfrak{r}}_\infty$ and the $R^{\chi,\emptyset,\mathfrak{r}}_\infty$-module $M^{\chi,\emptyset,\mathfrak{r}}_\infty$. In particular we conclude that $M^{1,\emptyset,\mathfrak{r}}_\infty$ is supported on every irreducible component of $\mathrm{Spec}(R^{1,\emptyset,\mathfrak{r}})'$.

As $M^{1,\emptyset,\mathfrak{r}}_\infty$ is a finite free S_∞-module, and $(R^{1,\emptyset,\mathfrak{r}}_\infty)^{\mathrm{red}}$ acts faithfully on $M^{1,\emptyset,\mathfrak{r}}_\infty$ (and is therefore finite and torsion free over S_∞), the map $\mathrm{Spec}(R^{1,\emptyset,\mathfrak{r}}_\infty) \to \mathrm{Spec}\, S_\infty$ is surjective and generalizing by [Sta13, Tag 080T]. It follows that we may pick some $\rho_{\mathrm{reg}} : G_F \to \mathrm{GSp}_4(\mathcal{O})$ whose corresponding point is in the support of $M^{1,\emptyset,\mathfrak{r}}/\mathfrak{p}_{\kappa_{\mathrm{reg}}}M^{1,\emptyset,\mathfrak{r}}$ and such that ρ and ρ_{reg} (or their corresponding points x and x_{reg}) lie on the same component of $R^{1,\emptyset,\mathfrak{r}}_\infty$; we write η for the generic point corresponding to this component. Similarly, we may pick some $\rho^\chi_{\mathrm{reg}} : G_F \to \mathrm{GSp}_4(\mathcal{O})$ whose corresponding point x^χ_{reg} is in the support of $M^{\chi,\emptyset,\mathfrak{r}}/\mathfrak{p}_{\kappa_{\mathrm{reg}}}M^{\chi,\emptyset,\mathfrak{r}}$.

By Proposition 7.13.11 below, we have

$$\dim_E(M^{1,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_{x_{\mathrm{reg}}}M^{1,\emptyset,\mathfrak{r}}_\infty)[1/p] = \dim_E S^{\emptyset,\mathfrak{r}}_{\kappa_{\mathrm{reg}},K^{p,\mathrm{Iw}}K_p(\emptyset)}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{r}}_{x_{\mathrm{reg}}}] = 8d_\rho,$$

and

$$\dim_E(M^{\chi,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_{x^\chi_{\mathrm{reg}}}M^{\chi,\emptyset,\mathfrak{r}}_\infty)[1/p] = \dim_E S^{\emptyset,\mathfrak{r}}_{\kappa_{\mathrm{reg}},K^{p,\mathrm{Iw}_1}K_p(\emptyset),\chi}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{r}}_{x^\chi_{\mathrm{reg}}}] = 8.$$

In addition, there are automorphic representations π_{reg}, π^χ_{reg} of $\mathrm{GSp}_4(\mathbf{A}_F)$ of weight κ_{reg} and central character $|\cdot|^2$ such that $\rho_{\pi_{\mathrm{reg}},p} \cong \rho_{\mathrm{reg}}$ and $\rho_{\pi^\chi_{\mathrm{reg}},p} \cong \rho^\chi_{\mathrm{reg}}$.

Applying Lemma 7.13.8 to $\mathfrak{p}_x$ and $\mathfrak{p}_{x_{\mathrm{reg}}}$ (which we may, by our assumptions on ρ, and by Theorem 2.7.2 for $\rho_{\pi_{\mathrm{reg}},p}$, together with Lemmas 7.1.3 and 7.3.18),we obtain

$$\dim_E(M^{1,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_x M^{1,\emptyset,\mathfrak{r}}_\infty)[1/p] = \dim_{(R^{1,\emptyset,\mathfrak{r}}_{\infty,\eta})'} M^{1,\emptyset,\mathfrak{r}}_{\infty,\eta} = \dim_E(M^{1,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_{x_{\mathrm{reg}}}M^{1,\emptyset,\mathfrak{r}}_\infty)[1/p] = 8d_\rho.$$

As

$$S^{\emptyset,\mathfrak{r}}_{\kappa_2,K^{p,\mathrm{Iw}},K_p(\emptyset)}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{r}}_x] = (M^{1,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_x M^{1,\emptyset,\mathfrak{r}}_\infty)[1/p]^\vee,$$

the theorem is proved for $I = \emptyset$.

Before we go on to the case that $\#I = 1$, we note that we may also apply Proposition 7.13.11 and Lemma 7.13.8 to $\mathfrak{p}_{x^\chi_{\mathrm{reg}}}$ and conclude that

$$\dim_{(R^{\chi,\emptyset,\mathfrak{r}}_{\infty,\eta^\chi})'} M^{\chi,\emptyset,\mathfrak{r}}_{\infty,\eta^\chi} = \dim_E(M^{\chi,\emptyset,\mathfrak{r}}_\infty/\mathfrak{p}_{x^\chi_{\mathrm{reg}}}M^{\chi,\emptyset,\mathfrak{r}}_\infty)[1/p] = 8.$$

By another application of Lemma 7.12.3 this implies that $d'_\eta = d_\rho$ (where d'_η is as in Lemma 7.12.3). Following Remark 7.12.6, we will apply this in the case $\#I = 1$ below.

Now consider the case that $\#I = 1$. We consider the automorphic representation π of Hypothesis 7.13.1. By assumption, for all finite places v of F the representation $\rho_{\pi,p}|_{G_{F_v}}$ is pure, and therefore determines a unique component of $R^{1,\emptyset,\mathfrak{r}}_\infty$,

which we denote by η_π. Arguing as above, we find that $d'_{\eta_\pi} = d_{\rho_{\pi,p}}$ (where d'_η is as in Lemma 7.12.3). Write $\tilde{\mathfrak{p}}_\pi^{I,\mathfrak{s}}$ for the height one prime ideal determined by $\rho_{\pi,p}$. Then by Proposition 7.13.11 we find that

$$\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_\pi^{I,\mathfrak{s}}] \geq 8d_{\rho_{\pi,p}}. \tag{7.13.10}$$

Again $M_\infty^{1,I,\mathfrak{s}}$ is a maximal Cohen–Macaulay $(R_\infty^{1,I,\mathfrak{s}})'$-module and so we may apply Lemmas 7.12.3 and 7.13.8 to the $R_\infty^{1,I,\mathfrak{s}}$-module $M_\infty^{1,I,\mathfrak{s}}$ and the $R_\infty^{\chi,I,\mathfrak{s}}$-module $M_\infty^{\chi,I,\mathfrak{s}}$. We find that

$$\begin{aligned}
\frac{1}{d_\rho}\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_x^{I,\mathfrak{s}}] &= \frac{1}{d_\rho}\dim_E(M_\infty^{1,I,\mathfrak{s}}/\tilde{\mathfrak{p}}_x^{I,\mathfrak{s}}M_\infty^{1,I,\mathfrak{s}})[1/p] \\
&= \frac{1}{d_\rho}\dim_{(R^{1,I,\mathfrak{s}}_{\infty,\eta})'} M^{1,I,\mathfrak{s}}_{\infty,\eta} \\
&= \dim_{(R^{\chi,I,\mathfrak{s}}_{\infty,\eta^\chi})'} M^{\chi,I,\mathfrak{s}}_{\infty,\eta^\chi} \\
&= \frac{1}{d_{\rho_{\pi,p}}}\dim_{(R^{1,I,\mathfrak{s}}_{\infty,\eta_\pi})'} M^{1,I,\mathfrak{s}}_{\infty,\eta_\pi} \\
&= \frac{1}{d_{\rho_{\pi,p}}}\dim_E(M_\infty^{1,I,\mathfrak{s}}/\tilde{\mathfrak{p}}_\pi^{I,\mathfrak{s}}M_\infty^{1,I,\mathfrak{s}})[1/p] \\
&= \frac{1}{d_{\rho_{\pi,p}}}\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_\pi^{I,\mathfrak{s}}].
\end{aligned}$$

It follows from (7.13.10) that

$$\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_x^{I,\mathfrak{s}}] \geq 8d_\rho.$$

On the other hand by Proposition 7.13.9, we have that

$$\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_x^{I,\mathfrak{s}}] \leq \dim_E S^{\emptyset,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(\emptyset)}[\tilde{\mathfrak{p}}_x^{\emptyset,\mathfrak{s}}] = 8d_\rho,$$

and so the theorem is proved. □

Proposition 7.13.11. *In the notation of the proof of Theorem 7.13.6, we have*

$$\begin{aligned}
&\dim_E S^{\emptyset,\mathfrak{s}}_{\kappa_{\mathrm{reg}},K^{p,\mathrm{Iw}}K_p(\emptyset)}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{s}}_{x_{\mathrm{reg}}}] = 8d_\rho, \\
&\dim_E S^{\emptyset,\mathfrak{s}}_{\kappa_{\mathrm{reg}},K^{p,\mathrm{Iw}_1}K_p(\emptyset),\chi}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{s}}_{x^\chi_{\mathrm{reg}}}] = 8, \\
&\dim_E S^{I,\mathfrak{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_\pi^{I,\mathfrak{s}}] \geq 8d_{\rho_{\pi,p}}.
\end{aligned}$$

In addition, there are automorphic representations π_{reg}, π^χ_{reg} of $\mathrm{GSp}_4(\mathbf{A}_F)$ of weight κ_{reg} and central character $|\cdot|^2$ such that $\rho_{\pi_{\mathrm{reg}},p} \cong \rho_{\mathrm{reg}}$ and $\rho_{\pi^\chi_{\mathrm{reg}},p} \cong \rho^\chi_{\mathrm{reg}}$.

Proof. By Theorem 6.6.5 and Theorem 3.10.1, $\dim_E S^{\emptyset,\mathfrak{s}}_{\kappa_{\mathrm{reg}},K^{p,\mathrm{Iw}}K_p(\emptyset)}[\tilde{\mathfrak{p}}^{\emptyset,\mathfrak{s}}_{x_{\mathrm{reg}}}]$ is equal to

$$\sum_\pi \dim_{\overline{E}}(\pi^\infty)^{K^{p,\mathrm{Iw}}K_p(\emptyset),\{U_{v,1}=\mathfrak{s}_v^{\mathrm{reg}},U_{v,2}=\alpha_v^{\mathrm{reg}}\beta_v^{\mathrm{reg}}\}_{v\in S_p}}$$

where the sum is over all the cuspidal automorphic representations π of weight κ_{reg} such that π has central character $|\cdot|^2$, π_v is holomorphic for all places $v|\infty$, and $\rho_{\pi,p} \cong \rho_{\mathrm{reg}}$; and we write $\alpha_v^{\mathrm{reg}}, \beta_v^{\mathrm{reg}}$ for the lifts of $\overline{\alpha}_v, \overline{\beta}_v$ determined by $\rho_{\mathrm{reg}}|_{G_{F_v}}$. In particular, note that we can take π_{reg} to be any of the automorphic representations π contributing to the sum.

Since ρ_{reg} is irreducible, such a π is of general type in the sense of [Art04] by Lemma 2.9.1, and therefore corresponds to an essentially self-dual regular algebraic

cuspidal automorphic representation Π of $\mathrm{GL}_4(\mathbf{A}_F)$. By strong multiplicity one for GL_4 [JS81], Π is uniquely determined by the condition that $\rho_{\Pi,p} \cong \rho_{\mathrm{reg}}$, so by Theorem 2.9.3 we see that we can rewrite the above sum as

$$\left(\sum_{\pi_{v_0} \in L(\rho_{\mathrm{reg}}|_{G_{F_{v_0}}})} \dim \pi_{v_0}^{\mathrm{Iw}_1(v_0)}\right) \prod_{v \in R} \left(\sum_{\pi_v \in L(\rho_{\mathrm{reg}}|_{G_{F_v}})} \dim \pi_v^{\mathrm{Iw}(v)}\right)$$

(note that at all places $v \notin R \cup S_p \cup \{v_0\}$, we are taking the space of hyperspecial invariants in an unramified representation, which is 1-dimensional; and at the places $v \in S_p$, the contribution is 1-dimensional by Propositions 2.4.24 and 2.4.26).

By Proposition 2.4.6, π_{v_0} is an irreducible unramified principal series representation; indeed, by the choice of v_0, $\rho_{\pi,p}|_{G_{F_{v_0}}}$ is unramified, and no two eigenvalues of $\rho_{\pi,p}|_{G_{F_{v_0}}}(\mathrm{Frob}_{v_0})$ can have ratio q_{v_0}. It follows from Propositions 2.4.3 and 2.4.4 that we have $\dim \pi_{v_0}^{\mathrm{Iw}_1(v_0)} = 8$. The claim then follows from Proposition 7.13.5 (which we can apply, because for each place $v \in R$, $\rho_{\mathrm{reg}}|_{G_{F_v}}$ is pure by Theorem 2.7.2 (4), and therefore the corresponding Weil–Deligne representation is generic by Lemma 7.1.3, so that the corresponding L-packet contains a generic representation by Proposition 2.4.22).

The statement for $\rho_{\mathrm{reg}}^{\chi}$ reduces in the same way to the claim that for each place $v \in R$, we have

$$\sum_{\pi_v \in L(\rho_{\mathrm{reg}}^{\chi}|_{G_{F_v}})} \dim \pi_v^{\mathrm{Iw}_1(v),\chi} = 1,$$

which follows from Proposition 2.4.28. Finally, in the case of $S^{I,\tilde{s}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}_\pi^{I,\tilde{s}}]$, the result follows as above, by computing the contribution of the automorphic representation π of Hypothesis 7.13.1 (note that it contributes by Theorem 3.10.1). □

Remark 7.13.12. In the course of the proof of Theorem 7.13.6, we showed that for the generic point η corresponding to ρ, the quantity d'_ρ of Lemma 7.12.3 is equal to d_ρ. It is presumably possible to go further following [Sho18], and to use our patched modules to show that for each $v \in R$ and each generic point η_v of $\operatorname{Spec} R_v^1$, if we write

$$Z(\operatorname{Spec} R_v^{1,\eta_v}/(\lambda)) = \overline{\eta}_v$$

then

$$Z(\operatorname{Spec}(R_v^{\chi})^{\mathrm{red}}/(\lambda)) = \sum_{\eta_v} d_{\eta_v} \overline{\eta}_v,$$

where d_{η_v} is as in Proposition 7.13.5.

8. Étale descent and the main modularity lifting theorem

8.1. **Introduction.** Our main goal is to remove the assumption $\#I \leq 1$ of Theorem 7.13.6 in order to eventually apply Theorem 6.5.8 with $I = S_p$, and from this conclude that we have constructed classical automorphic representations. The starting point is to consider the spaces of p-adic automorphic forms considered in Theorem 7.13.6 for *both* $\#I = 1$ and $\#I = 0$. By studying the way in which these spaces are related, we will be able to (inductively) determine precise linear combinations of such forms which belong to spaces of p-adic automorphic forms for larger $\#I$. Our argument uses the doubling results of §5, the analytic continuation

results of §6, and étale descent. Finally, we apply solvable base change to prove our main modularity lifting theorem.

We briefly indicate some of the main features of our argument. As we mentioned in the introduction, the analytic continuation arguments that we are using here are analogous to those used for Hilbert modular forms of weight at least two, rather than those of weight one – in particular, there is no "gluing" of the kind used in [BT99], and we are simply analytically continuing a single form at a time (using the method of Kassaei series [Kas06]). This part of the argument is quite standard, although we have to take some care to show that the regions that we have analytically continued to are large enough. For this reason, we ignore the issues of analytic continuation in this introduction.

We show that the conclusion of Theorem 7.13.6 holds for all I by induction on $\#I$. The key step is to go from $\#I \leq 1$ to $\#I \leq 2$; indeed, the general inductive step considers two places v_1, v_2 dividing p, and essentially ignores the other places above p, so for the purpose of exposition we assume that $S_p = \{v_1, v_2\}$. Write α_i, β_i for $\alpha_{v_i}, \beta_{v_i}$, $i = 1, 2$. We denote the various spaces of forms considered in Theorem 7.13.6 with $I = \emptyset$ by V_{α_1,α_2}, V_{β_1,α_2}, V_{α_1,β_2}, V_{β_1,β_2} (so that for example on V_{α_1,α_2} the eigenvalue of $U_{v_1,1}$ is α_1 and the eigenvalue of $U_{v_2,1}$ is α_2). Each of these spaces has dimension $d := 8d_\rho$, and considering the action of $U_{v_1,1}$ and $U_{v_2,1}$, we see that these spaces together span a $4d$-dimensional space of p-adic modular forms of Iwahori level.

We expect that this space contains a d-dimensional subspace of p-adic modular forms which descend to Klingen level (and are suitably overconvergent in both the v_1 and v_2 directions). The difficulty (even if $d = 1$) is in identifying this subspace; recall that there is no obvious relationship between the Hecke eigenvalues and Fourier coefficients. However, we also have the spaces of forms for $I = \{v_1\}$ and $I = \{v_2\}$, which we denote by $V_{\alpha_1,\alpha_2+\beta_2}$, $V_{\beta_1,\alpha_2+\beta_2}$, $V_{\alpha_1+\beta_1,\alpha_2}$, $V_{\alpha_1+\beta_1,\beta_2}$, where for example the forms in $V_{\alpha_1,\alpha_2+\beta_2}$ have Klingen level at v_2 (and are highly overconvergent in the v_2 direction), and are $U_{\mathrm{Kli}(v_2),1}$-eigenforms with eigenvalue $\alpha_2 + \beta_2$. Again, all of these spaces has dimension d by Theorem 7.13.6.

Now, the relations between the Hecke operators at Klingen and Iwahori levels (more precisely, Lemma 4.5.17) imply that we have a map

$$(U_{v_1,1} - \beta_1) : V_{\alpha_1+\beta_1,\alpha_2} \to V_{\alpha_1,\alpha_2}.$$

Furthermore, this map is injective by Proposition 7.13.9 (that is, by our main doubling results), and since the source and target both have dimension d, this map is in fact an isomorphism. Similarly, we have an isomorphism

$$(U_{v_1,1} - \beta_1) : V_{\alpha_1+\beta_1,\beta_2} \xrightarrow{\sim} V_{\alpha_1,\beta_2}$$

and thus an isomorphism of $2d$-dimensional spaces

$$(U_{v_1,1} - \beta_1) : V_{\alpha_1+\beta_1,\alpha_2} \oplus V_{\alpha_1+\beta_1,\beta_2} \xrightarrow{\sim} V_{\alpha_1,\alpha_2} \oplus V_{\alpha_1,\beta_2}. \tag{8.1.1}$$

By pulling back from Iwahori to Klingen level, we can think of $V_{\alpha_1,\alpha_2+\beta_2}$ as a d-dimensional subspace of the target of (8.1.1). The inverse image of this space in the source of (8.1.1) is the d-dimensional space of forms that we are seeking; considered as living on the right hand side of (8.1.1), it comes from Klingen level at v_2, and on the left hand side of (8.1.1), it comes from Klingen level at v_1. We make this precise using an argument with étale descent.

8.2. **Étale descent.** In this section we carry out the argument explained above, showing that the conclusion of Theorem 7.13.6 holds for all I by induction on $\#I$ (in fact, we show slightly more, keeping track of the overconvergence of our p-adic modular forms). Recall that by definition for each choice of $I, \mathfrak{r}$ we have

$$S^{I,\mathfrak{r}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)} = (H^0(\mathfrak{X}^{I,G_1}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D))_{\tilde{\mathfrak{m}}^{I,\mathfrak{r}}}) \otimes_{\mathcal{O}} E[\{U_{v,0}-1\}_{v\in S_p}, \\ \{U_{v,0}-q_v^{-2}\}_{v\in R}, \{T_{v,0}-q_v^{-2}\}_{v\notin S_p\cup R\cup\{v_0\}}]$$

The maximal ideal $\tilde{\mathfrak{m}}^{I,\mathfrak{r}}$ of the Hecke algebra is defined in equation (7.8.4). It contains an ordinary projector. We have given ourselves (see the beginning of §7.13) a Galois representation ρ satisfying Hypothesis 7.13.1. We want to prove that it is modular. Associated to this representation is a point x on the deformation space of $\bar{\rho}$ and an ideal $\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x$ (see equation (7.13.3)) of the Hecke algebra contained in $\tilde{\mathfrak{m}}^{I,\mathfrak{r}}$ whose definition we recall here for convenience. It is the ideal of the Hecke algebra $\widetilde{\mathbf{T}}^I$ given by

$$\left(\bigotimes_{v\notin S_p\cup R\cup\{v_0\}} \mathcal{O}[\mathrm{GSp}_4(F_v)/\!/\,\mathrm{GSp}_4(\mathcal{O}_{F_v})]\right)[\{U_{v,0}, U_{\mathrm{Kli}(v),1}, U_{v,2}\}_{v\in I}, \{U_{v,0}, U_{v,1}, U_{v,2}\}_{v\in I^c}]$$

which is generated by:

- the coefficients of $\det(X-\rho(\mathrm{Frob}_v)) - Q_v(X)$ for each $v \notin S_p \cup R \cup \{v_0\}$, and
- $\{U_{v,0}-1, U_{v,2}-\alpha_v\beta_v\}_{v\in S_p}, \{U_{\mathrm{Kli}(v),1}-\alpha_v-\beta_v\}_{v\in I}, \{U_{v,1}-\mathfrak{r}_v\}_{v\in I^c}$, where, for $v \in S_p$, α_v, β_v are determined by $\rho|_{G_{F_v}}$ as in Definition 7.3.1.

Recall that $\mathcal{X}^{I,G_1}_{K^{p,\mathrm{Iw}}K_p(I)}$ is the analytic adic space over $\mathbf{C}_p$ associated to $\mathfrak{X}^{I,G_1}_{K^{p,\mathrm{Iw}}K_p(I)}$. By definition, we have:

$$S^{I,\mathfrak{r}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x] \otimes_E \mathbf{C}_p = e(U^I)H^0(\mathcal{X}^{I,G_1}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D))[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x].$$

We may also introduce overconvergent versions of these spaces. Recall that we defined the dagger space $\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(I)}$ in (6.5.5) (whose associated rigid analytic space is $\mathcal{X}^{G_1,I}_{K^{p,\mathrm{Iw}}K_p(I)}$).

There is a natural injective restriction map:

$$e(U^I)H^0(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D)) \to e(U^I)H^0(\mathcal{X}^{G_1,I}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D)).$$

Let

$$S^{I,\mathfrak{r},\dagger}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)} = e(U^I)H^0(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D)) \cap S^{I,\mathfrak{r}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)} \otimes \mathbf{C}_p$$

where the intersection is taken inside $e(U^I)H^0(\mathcal{X}^{G_1,I}_{K^{p,\mathrm{Iw}}K_p(I)}, \omega^2(-D))$.

Theorem 8.2.1. *Assume that ρ satisfies Hypothesis 7.13.1. Then for any $I \subset S_p$ and choice of $\mathfrak{r}$, we have*

$$\dim_E S^{I,\mathfrak{r}}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x] = \dim_{\mathbf{C}_p} S^{I,\mathfrak{r},\dagger}_{\kappa_2,K^{p,\mathrm{Iw}}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x] = 8d_\rho.$$

Before proving this theorem, we record the following important corollary.

Corollary 8.2.2. *Suppose that ρ satisfies Hypothesis 7.13.1. Then ρ is modular. More precisely, there is an ordinary automorphic representation π' of* $\mathrm{GSp}_4(\mathbf{A}_F)$

of parallel weight 2 *and central character* $|\cdot|^2$*, with* $\rho_{\pi',p} \cong \rho$*, and for every finite place* v *of* F *we have*

$$\mathrm{WD}(\rho|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_v \otimes |\nu|^{-3/2}).$$

Proof. The existence of π' with $\rho_{\pi',p} \cong \rho$ is immediate from Theorem 8.2.1, taking $I = S_p$, together with Theorem 6.5.8 and Theorem 3.10.1. By Corollary 7.9.6 we have

$$\mathrm{WD}(\rho|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}$$

at all finite places v of F, so we need only prove that the monodromy operators agree at the places $v \in R$. Since $\rho|_{G_{F_v}}$ is pure by assumption, it follows from Lemma 2.5.1 that it suffices to prove, in the notation of Section 2.3, that $n(\rho|_{G_{F_v}}) \leq n(\pi'_v)$.

Now, if π_v is any irreducible admissible representation of $\mathrm{GSp}_4(F_v)$, then an examination of [RS07b, Table A.15] (noting that the column there headed "a" records $n(\pi_v)$) shows that:

- $n(\pi_v) \geq 1$ if and only if $(\pi_v)^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} = 0$.
- $n(\pi_v) \geq 2$ if and only if $(\pi_v)^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} = (\pi_v)^{\mathrm{Par}(v)} = 0$.
- $n(\pi_v) = 3$ if and only if $(\pi_v)^{\mathrm{Kli}(v)} = (\pi_v)^{\mathrm{Si}(v)} = 0$.

Suppose that $n(\rho|_{G_{F_v}}) = 1$, so that we need to show that $(\pi'_v)^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} = 0$. Suppose for the sake of contradiction that $(\pi'_v)^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \neq 0$; then by Hida theory (more precisely, by Theorem 7.9.4 and its proof), the Galois representation $\rho_{\pi',p}$ is a p-adic limit of Galois representations $\rho_{\pi'',p}$, where π'' has regular weight and satisfies $(\pi''_v)^{\mathrm{GSp}_4(\mathcal{O}_{F_v})} \neq 0$. In particular, by Theorem 2.7.2, $n(\rho_{\pi'',p}|_{G_{F_v}}) = n(\pi''_v) = 0$. By the semicontinuity of the rank of the nilpotent operator N in such a family, it follows that $n(\rho|_{G_{F_v}}) = 0$, a contradiction. We leave the (very similar) arguments in the cases $n(\rho|_{G_{F_v}}) = 2, 3$ to the reader. □

Proof of Theorem 8.2.1. We prove this by induction on $\#I$. The result is true for $I = \emptyset$ and $\#I = 1$ by Theorem 7.13.6 and Theorem 6.6.4 (ordinary implies overconvergent if $\#I \leq 1$). For any I, the restriction map

$$S^{I,\mathfrak{r},\dagger}_{\kappa_2,K^p,\mathrm{Iw}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x] \to S^{I,\mathfrak{r}}_{\kappa_2,K^p,\mathrm{Iw}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x] \otimes_E \mathbf{C}_p$$

is injective, while by Proposition 7.13.9 (and a simple induction) we see that for any I, the dimension of $S^{I,\mathfrak{r}}_{\kappa_2,K^p,\mathrm{Iw}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x]$ is at most $8d_\rho$. It therefore suffices to show that $S^{I,\mathfrak{r},\dagger}_{\kappa_2,K^p,\mathrm{Iw}K_p(I)}[\tilde{\mathfrak{p}}^{I,\mathfrak{r}}_x]$ has dimension at least $8d_\rho$. We may assume that $\#I \geq 2$, and hence we may write I as a disjoint union $J \cup \{v_1, v_2\}$ for two primes $v_i|p$. We fix the choice of $\mathfrak{r}$ at all primes in $J \cup I^c$.

By the inductive hypothesis applied to J, for each choice of $\mathfrak{r}$ at v_1 and v_2 the corresponding eigenspace $S^{J,\mathfrak{r},\dagger}_{\kappa_2,K^p,\mathrm{Iw}K_p(J)}[\tilde{\mathfrak{p}}^{J,\mathfrak{r}}_x]$ is $8d_\rho$-dimensional, and we denote these eigenspaces by V_{α_1,α_2}, V_{β_1,α_2}, V_{α_1,β_2}, V_{β_1,β_2} (so that for example on V_{α_1,α_2} the eigenvalue of $U_{v_1,1}$ is α_1 and the eigenvalue of $U_{v_2,1}$ is α_2). Considering the action of $U_{v_1,1}$ and $U_{v_2,1}$, we see that these spaces span a $4 \times 8d_\rho = 32d_\rho$-dimensional subspace V of $H^0(\mathcal{X}^{J,G_1}_{K^p,\mathrm{Iw}K_p(J)}, \omega^2(-D))$.

By the inductive hypothesis applied to $J \cup \{v_1\}$ and the two possible choices of $\mathfrak{r}$ at v_2 (the choice at v_1 is irrelevant), we see that the eigenspaces $S^{J\cup\{v_1\},\mathfrak{r},\dagger}_{\kappa_2,K^p,\mathrm{Iw}K_p(J\cup\{v_1\})}[\tilde{\mathfrak{p}}^{J\cup\{v_1\},\mathfrak{r}}_x]$ are both $8d_\rho$-dimensional, and we denote the corresponding spaces by $V_{\alpha_1+\beta_1,\alpha_2}$ and $V_{\alpha_1+\beta_1,\beta_2}$. Similarly, the inductive hypothesis applied to $J \cup \{v_2\}$ yields $8d_\rho$-dimensional spaces $V_{\alpha_1,\alpha_2+\beta_2}$ and $V_{\beta_1,\alpha_2+\beta_2}$.

Recall that our goal is to construct an $8d_\rho$-dimensional space of eigenforms $V_{\alpha_1+\beta_1,\alpha_2+\beta_2}$ which are eigenforms for the operators $U_{\mathrm{Kli}(v_1),1}$ and $U_{\mathrm{Kli}(v_2),1}$ (and for the Hecke operators at all the other places), and which lie in

$$H^0(\mathcal{X}^{G_1,\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(I)},\omega^2(-D)).$$

We will combine the analytic continuation results of §6 with a descent argument to prove the existence of the sought-after eigenforms.

We need to introduce some notation in order to be able to describe the adic spaces we are working with. Recall that $\mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}$ is the analytic space associated to $X_{K^{p,\mathrm{Iw}}K_p(I)}$. For each $v \in I$, H_v refers to the quasi-finite subgroup (of order p over the interior of the moduli space) related to the Klingen level structure, and for each $v \in I^c$, $L_v \supset H_v$ refers to the quasi-finite (maximally isotropic rank p^2 over the interior of the moduli space) subgroup corresponding to the Iwahori level structure.

For any tuple $(\epsilon_v) \in [0,1]^I \times [0,2]^{I^c}$ we defined an analytic adic space $\mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}((\epsilon_v)_{v\in S_p})$ which is the open subspace of $\mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}$ where:

(1) If $v \in I$, the degree of the subgroup H_v, which takes values in $[0,1]$, is greater or equal than $1-\epsilon_v$.
(2) If $v \in I^c$, the degree of the subgroup L_v of rank p^2, which takes values in $[0,2]$, is greater or equal than $2-\epsilon_v$. Note that we have $\deg(L_v) = \deg(H_v) + \deg(L_v/H_v)$.

It will be convenient to adopt the following notation in this proof (note that I is fixed). We write (cf. (6.5.5))

$$\mathcal{X}^{\mathrm{mult}} = \mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}((0)_{v\in S_p}),$$

$$\mathcal{X}^{\mathrm{mult},\dagger} = \mathcal{X}^{\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(I)} = \lim_{\epsilon_v\to 0^+} \mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}((\epsilon_v)_{v\in S_p}),$$

and $\mathcal{X}^{\mathrm{mult},\ddagger}$ for the dagger space

$$\mathcal{X}^{\mathrm{mult},\ddagger} := \lim_{\epsilon_v\to 0^+} \mathcal{X}_{K^{p,\mathrm{Iw}}K_p(I)}((\epsilon_v)_{v\in S_p\setminus\{v_1,v_2\}}),$$

where we take the limit over all primes *except* v_1 and v_2. It follows that:

$$\mathcal{X}^{\mathrm{mult},\dagger} = \lim_{\epsilon_{v_1},\epsilon_{v_2}\to 0^+} \mathcal{X}^{\mathrm{mult},\ddagger}(\epsilon_{v_1},\epsilon_{v_2}),$$

and there are maps of locally ringed spaces $\mathcal{X}^{\mathrm{mult}} \to \mathcal{X}^{\mathrm{mult},\dagger} \to \mathcal{X}^{\mathrm{mult},\ddagger}$. By adding the subscript $\mathrm{Iw}(v_1)$ (or $\mathrm{Iw}(v_2)$, or $\mathrm{Iw}(v_1,v_2)$) to $\mathcal{X}^{\mathrm{mult}}$, $\mathcal{X}^{\mathrm{mult},\ddagger}$, $\mathcal{X}^{\mathrm{mult},\dagger}$ we mean the space where one has now added an Iwahori level structure at v_1 (or v_2, or v_1 and v_2) to the relevant space. For $i=1,2$ we write $d_i^H = \deg H_{v_i}$, $d_i^L = \deg L_{v_i}$, whenever these quantities are defined. We will adorn $\mathcal{X}^{\mathrm{mult}}$ and $\mathcal{X}^{\mathrm{mult},\ddagger}$ with superscripts indicating the regions (which will typically strictly contain $\mathcal{X}^{\mathrm{mult}}$ and $\mathcal{X}^{\mathrm{mult},\ddagger}$) where various inequalities hold.

Returning to the spaces we defined above, we have

$$\begin{aligned} V_{\mathfrak{s}_1,\mathfrak{s}_2} &\subset H^0(\mathcal{X}^{\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(J)},\omega^2(-D)),\\ V_{\mathfrak{s}_1+\mathfrak{s}_1',\mathfrak{s}_2} &\subset H^0(\mathcal{X}^{\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(J\cup\{v_1\})},\omega^2(-D)),\\ V_{\mathfrak{s}_1,\mathfrak{s}_2+\mathfrak{s}_2'} &\subset H^0(\mathcal{X}^{\mathrm{mult},\dagger}_{K^{p,\mathrm{Iw}}K_p(J\cup\{v_2\})},\omega^2(-D)). \end{aligned}$$

Lemma 8.2.3. *The elements of $V_{\mathfrak{s}_1,\mathfrak{s}_2}$ extend to $\mathcal{X}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon,d_1^L>1,d_2^L>1}_{\mathrm{Iw}(v_1,v_2)}$ for some $\epsilon>0$. Similarly, the elements of $V_{\mathfrak{s}_1+\mathfrak{s}_1',\mathfrak{s}_2}$ and $V_{\mathfrak{s}_1,\mathfrak{s}_2+\mathfrak{s}_2'}$ extend to the*

spaces $\mathcal{X}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H>1-\epsilon,d_2^L>1}_{\mathrm{Iw}(v_2)}$ *and* $\mathcal{X}^{\mathrm{mult},\ddagger,d_1^H>1-\epsilon,d_1^L>1,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_1)}$ *respectively for some* $\epsilon>0$.

Proof. This follows from Lemma 6.5.18 (taking I there to be J, $J\cup\{v_2\}$ and $J\cup\{v_1\}$ respectively). Note that our forms are ordinary for $U_{w,1}$ for the appropriate w, and therefore of finite slope for these operators. □

By Koecher's principle, all of our cohomology groups may be replaced by the cohomology of the corresponding open spaces of "good reduction" $\mathcal{Y}^{\mathrm{mult}}\subset\mathcal{X}^{\mathrm{mult}}$, $\mathcal{Y}^{\mathrm{mult},\ddagger}\subset\mathcal{X}^{\mathrm{mult},\ddagger}$, and $\mathcal{Y}^{\mathrm{mult},\dagger}\subset\mathcal{X}^{\mathrm{mult},\dagger}$ respectively. (Since the sheaf ω^2 is pulled back from the minimal compactification, the form of Koecher's principle we are using is just the following statement: if $\mathfrak{X}$ is a normal formal scheme, $\mathfrak{Y}\subseteq\mathfrak{X}$ is an open formal subscheme whose complement is codimension ≥ 2, and $\mathcal{L}/\mathfrak{X}$ is a line bundle, then $H^0(\mathfrak{X},\mathcal{L})=H^0(\mathfrak{Y},\mathcal{L})$. That the boundary in the minimal compactification does indeed have codimension ≥ 2 follows from an analysis of the blowup in the boundary charts.) We now restrict to these spaces to avoid minor technical issues related to the boundary. In particular we will want to use that forgetting the level structure induces finite étale maps between our spaces. The reader will check easily that the forms we construct are indeed cuspidal because they are obtained by "descent" of cuspidal forms.

For any $\epsilon>0$ and for $i=1,2$ there is a finite étale map:

$$q_{v_i}:\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_i)}\to\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}.$$

There is a corresponding fibre product map (still finite étale):

$$q_{v_1}:\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_1,v_2)}\to\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_2)}$$

(and similarly for q_{v_2}).

We now somewhat abusively also write $V_{\alpha_1+\beta_1,\alpha_2}$ instead of $q_{v_1}^*V_{\alpha_1+\beta_1,\alpha_2}$. We claim that the action of $(U_{v_1,1}-\beta_1)$ induces an isomorphism

$$(U_{v_1,1}-\beta_1):V_{\alpha_1+\beta_1,\alpha_2}\xrightarrow{\sim}V_{\alpha_1,\alpha_2}. \tag{8.2.4}$$

To see this, note that $(U_{v_1,1}-\beta_1)$ is injective by Proposition 7.13.9, and both spaces have the same dimension.

In the same way, we have an isomorphism $(U_{v_1,1}-\alpha_1):V_{\alpha_1+\beta_1,\alpha_2}\xrightarrow{\sim}V_{\beta_1,\alpha_2}$, so we see that in fact the span of $V_{\alpha_1+\beta_1,\alpha_2}$ and $U_{v_1,1}V_{\alpha_1+\beta_1,\alpha_2}$ is exactly $V_{\alpha_1,\alpha_2}\oplus V_{\beta_1,\alpha_2}$. It follows from Lemma 8.2.3 that the forms in $V_{\alpha_1,\alpha_2}\oplus V_{\beta_1,\alpha_2}$ extend to $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon,d_2^L>1}_{\mathrm{Iw}(v_1,v_2)}$ for some $\epsilon>0$.

Set

$$U_\epsilon:=\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon,d_2^L>1}_{\mathrm{Iw}(v_1,v_2)}.$$

The map q_{v_2} restricts to an étale map:

$$q_{v_2}:U_\epsilon\to\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_1)}.$$

We claim that for ϵ sufficiently small, the restriction of q_{v_2} to U_ϵ is surjective. Note that a pre-image of a point in $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}_{\mathrm{Iw}(v_1,v_2)}$ (without any condition on d_2^L) corresponds to a choice of $L=L_{v_2}$, which is determined by a line in $H^\perp/H\subset A[v_2]/H$ for $H=H_{v_2}$. We need to show that $\deg(L)>1$ for at least one such L.

Let us first assume that $\deg(H) = 1$. Then we can choose any line $C \subset H^{\perp}/H$ with $\deg(C) > 0$ (such a C exists as $H^{\perp}/H$ is not étale) and the corresponding L has $\deg(L) = \deg(H) + \deg(C) > 1$.

We pass from $\deg(H) = 1$ to $\deg(H) > 1 - \varepsilon$ by a continuity argument. The function which sends a rank one point $x \in \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}$ to the maximum of $\deg(L)$ is continuous. It follows that for ε sufficiently small, we can ensure the existence of a subgroup L such that $\deg(L) > 1$.

Consider the corresponding descent diagram:

$$U_\epsilon \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} U_\epsilon \underset{q_{v_2,2}}{\overset{q_{v_2,1}}{\rightrightarrows}} U_\epsilon \xrightarrow{q_{v_2}} \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}.$$

Lemma 8.2.5. *After possibly further shrinking $\epsilon > 0$, any element of $V_{\alpha_1,\alpha_2+\beta_2}$ descends to $\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}$.*

Proof. Any element of $V_{\alpha_1,\alpha_2+\beta_2}$ tautologically satisfies descent over the (smaller) space $\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1} \subset U_\epsilon$ to

$$q_{v_2}(\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1}) = \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon}$$

since it is (by Lemma 8.2.3) obtained simply by pulling back a form on this space under q_{v_2}. Therefore, we deduce that for any element $G \in V_{\alpha_1,\alpha_2+\beta_2}$, we have that $q_{v_2,1}^* G = q_{v_2,2}^* G$ on

$$\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1} \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} \mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1}.$$

The point is now to show that each connected component of

$$U_\epsilon \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} U_\epsilon$$

intersects

$$\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1} \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} \mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_1^L > 1, d_2^H \geq 1-\epsilon, d_2^L > 1}$$

so that we have that $q_{v_2,1}^* G = q_{v_2,2}^* G$ on $U_\epsilon \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} U_\epsilon$ and can perform the descent of G.

It follows from [Poi08, Thm. 2] that after possibly further shrinking ϵ, there is a surjective map

$$\pi_0(U_0 \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger d_1^H = d_2^H = 1}} U_0) \to \pi_0(U_\epsilon \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} U_\epsilon). \tag{8.2.6}$$

We need to see that $q_{v_2,1}^* G = q_{v_2,2}^* G$ on

$$U_\epsilon \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H \geq 1-\epsilon, d_2^H \geq 1-\epsilon}} U_\epsilon.$$

By (8.2.6), it is enough to show $q_{v_2,1}^* G = q_{v_2,2}^* G$ on the subspace

$$U_0 \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H = d_2^H = 1}} U_0.$$

As discussed above, this identity holds over the region

$$\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H = d_2^H = 1, d_1^L > 1, d_2^L > 1} \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H = d_2^H = 1}} \mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H = d_2^H = 1, d_1^L > 1, d_2^L > 1}$$

by definition. It therefore suffices to show that this region intersects all connected components of $U_0 \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1}} U_0$.

Accordingly, it is enough to show that every connected component of $U_0 \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1}} U_0$ contains a point which is non-ordinary at v_1. Indeed, if such a point had $d_1^L = 1$, then we would have $\deg(H_{v_1}) = \deg(L_{v_1}) = 1$, which implies that $\deg(L_{v_1}/H_{v_1}) = 0$, so L_{v_1}/H_{v_1} is étale, and the point is ordinary at v_1, a contradiction.

We will prove in Corollary 8.2.9 below that any connected component of either of the spaces

$$\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1},\ \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}2}$$

contains a point which is non-ordinary at v_1. Recall that the superscripts $=_{v_2} 1$ and $=_{v_2} 2$ respectively mean the rank 1 and the ordinary locus at v_2.

Now we observe that the maps $U_0^{=_{v_2}1} \to \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1}$ and $U_0^{=_{v_2}2} \to \mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}2}$ are both finite étale. It follows from Lemma 8.2.7 below that any connected component of any of the spaces $U_0^{=_{v_2}1}$, $U_0^{=_{v_2}2}$, $U_0^{=_{v_2}1} \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1}} U_0^{=_{v_2}1}$ or $U_0^{=_{v_2}2} \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1}} U_0^{=_{v_2}2}$ contains a point which is non-ordinary at v_1. It finally follows that any component of $U_0 \times_{\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1}} U_0$ contains a point which is non-ordinary at v_1, as required. □

We can now complete the proof of Theorem 8.2.1. Consider the diagram:

$$\begin{tikzcd}
\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon} \arrow[r, "q_{v_1}"] \arrow[d, "q_{v_2}"] & \mathcal{Y}_{\mathrm{Iw}(v_2)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon} \arrow[d, "q_{v_2}"] \\
\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon} \arrow[r, "q_{v_1}"] & \mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}
\end{tikzcd}$$

By Lemma 8.2.5, we have proved that all elements of our spaces $V_{\alpha_1+\beta_1,\alpha_2}$ and $V_{\alpha_1+\beta_1,\beta_2}$ are sections on the whole $\mathcal{Y}_{\mathrm{Iw}(v_2)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$ and that all elements of our spaces $V_{\alpha_1,\alpha_2+\beta_2}$ and $V_{\beta_1,\alpha_2+\beta_2}$ are sections on the whole $\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$. We can pull back these sections to $\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$.

The isomorphism (8.2.4) and the similar isomorphism $(U_{v_1,1}-\beta_1) : V_{\alpha_1+\beta_1,\beta_2} \xrightarrow{\sim} V_{\alpha_1,\beta_2}$ induce an isomorphism

$$(U_{v_1,1} - \beta_1) : V_{\alpha_1+\beta_1,\alpha_2} \oplus V_{\alpha_1+\beta_1,\beta_2} \xrightarrow{\sim} V_{\alpha_1,\alpha_2} \oplus V_{\alpha_1,\beta_2},$$

and we define $V_{\alpha_1+\beta_1,\alpha_2+\beta_2}$ to be the preimage of $V_{\alpha_1,\alpha_2+\beta_2} \subset V_{\alpha_1,\alpha_2} \oplus V_{\alpha_1,\beta_2}$ under this isomorphism. This is an $8d_\rho$-dimensional space of eigenforms with the appropriate eigenvalues, so we only need to check that all of the elements of $V_{\alpha_1+\beta_1,\alpha_2+\beta_2}$ descend to $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$.

Consider an element F of this space. By definition, F has the property that $(U_{v_1,1}-\beta_1)F$ on $\mathcal{Y}_{\mathrm{Iw}(v_1,v_2)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$ is pulled back from $\mathcal{Y}_{\mathrm{Iw}(v_1)}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$ via $q_{v_2}^*$. Let $G = \deg(q_{v_2})^{-1} q_{v_2,*}F$ be the trace of F to $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$. The form F

comes via pullback from $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H\geq 1-\epsilon,d_2^H\geq 1-\epsilon}$ if and only if $F = q_{v_2}^* G$. Since the trace map at v_2 commutes with $U_{v_1,1}$ (for the usual reasons, ultimately coming down to Serre–Tate theory and the product structure on the p-divisible group), we deduce (since q_{v_1} is surjective) that $(U_{v_1,1} - \beta_1)(q_{v_2}^* G - F) = 0$, so that $F = q_{v_2}^* G$ (because $(U_{v_1,1} - \beta_1)$ is injective) as required. □

We conclude this section with some lemmas that were used above. We first record the following easy lemma:

Lemma 8.2.7. *If $S \to T$ is a finite étale map of adic spaces of finite type over a field, then the image of any connected component of S is a connected component of T.*

Proof. Since S and T are of finite type, they have only finitely many connected components. In particular the connected components of S and T are precisely the connected subsets of S and T which are both open and closed. Since finite étale morphisms are both open and closed [Hub96, Lem. 1.4.5, Prop. 1.7.8], the result is immediate. □

Next we have the following lemma and its corollary:

Lemma 8.2.8. *Any connected component of $Y^{I,=_{v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ contains a point in $Y^{I,=_{\{v_1,v_2\}}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$, and any connected component of $Y^{I,=_{v_2}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$ contains a point in $Y^{I,=_{v_1}1,=_{v_2}2}_{K_p(I)K^p,1}$.*

Corollary 8.2.9. *Any connected component of either of the spaces*

$$\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1}_{\mathrm{Iw}(v_1)},\ \mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}2}_{\mathrm{Iw}(v_1)}$$

contains a point which is non-ordinary at v_1.

Proof. The map $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1}_{\mathrm{Iw}(v_1)} \to \mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1}$ is finite étale, so it suffices to prove the claims for $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}1}$ and $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1,=_{v_2}2}$. Also, the map $\mathcal{Y}^{\mathrm{mult},\ddagger,d_1^H=d_2^H=1} \to \mathcal{Y}^{\mathrm{mult}}$ induces an isomorphism of π_0's, because both spaces have the same rank one points and any higher rank point admits a generalization to a rank one point. Thus $\mathcal{Y}^{\mathrm{mult},=_{v_2}1}$ is the tube of $Y^{I,=_{v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ and $\mathcal{Y}^{\mathrm{mult},=_{v_2}2}$ is the tube of $Y^{I,=_{v_2}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$. Since all these spaces are smooth, the tube of a connected component in $Y^{I,=_{v_2}i}_{K^{p,\mathrm{Iw}}K_p(I),1}$ is connected for $i = 1, 2$. But now by Lemma 8.2.8 these components contain points which have rank one at v_1 and hence are not ordinary at v_1. □

The rest of this section is devoted to proving Lemma 8.2.8. This statement is a very special case of a general expectation that "all possible specializations between EKOR strata are realized." Unfortunately, as far as we are aware, this exact statement does not yet appear in the literature, but we will explain how it can be deduced from what is available using standard techniques. This will necessitate a small digression into the theory of stratifications of special fibres of Shimura varieties.

To aid the reader's understanding, we first recall a general strategy for producing specializations between strata: first one produces a specialization to a point of a very special stratum, and then one uses deformation theory at that special point

to "go back up" to the desired stratum. For achieving the first step, there is also a standard strategy: if one can show that open strata are (quasi)-affine, while the closures of strata are proper, then it follows that any component of any stratum must specialize to a point of a zero dimensional stratum. This argument becomes a bit more complicated for non compact Shimura varieties, where one must study the extension of the stratification to the boundary of the minimal compactification. In order to use results readily available in the literature, we will carry out the first step at spherical level, then carry out the second step at Iwahori level, and finally explain how this implies the result that we want at level $K_p(I)$.

First we consider the Ekedahl–Oort stratification at spherical level, see for instance [VW13]. Let $K_p = \prod_{v|p} \mathrm{GSp}_4(\mathcal{O}_{F_v})$. Then $Y_{K^p, Iw K_p, 1}$ has an Ekedahl–Oort stratification into $4^{[F:\mathbf{Q}]}$ strata, according to the four possibilities for each of the finite flat group schemes $\mathcal{G}_w[p]$ at geometric points. Let $G_{1,1} = E[p]$ for E a supersingular elliptic curve. Then these four possibilities are:

- Ordinary: $\mathcal{G}_w[p] \simeq \mu_p^2 \times (\mathbf{Z}_p/p\mathbf{Z}_p)^2$
- p-rank 1: $\mathcal{G}_w[p] \simeq \mu_p \times \mathbf{Z}_p/p\mathbf{Z}_p \times G_{1,1}$
- Supergeneral: $\mathcal{G}_w[p]$ is connected-connected, but not isomorphic to $G_{1,1}^2$.
- Superspecial: $\mathcal{G}_w[p] \simeq G_{1,1}^2$.

This stratification refines the p-rank stratification, with the last two cases corresponding to $\mathcal{G}_w$ having p-rank 0. We call a point of $Y_{K^p, Iw K_p, 1}$ superspecial if $\mathcal{G}_w[p]$ is superspecial for all $w|p$. This is the unique zero dimensional stratum.

Lemma 8.2.10. *Let $J \subseteq S_p$. Each irreducible component of $Y^{=_{J^c}2, =_J 1}_{K^p, \mathrm{Iw} K_p, 1}$ contains a point of $Y^{=_{S_p}0}_{K^p, \mathrm{Iw} K_p, 1}$ in its closure.*

Proof. It is shown in [Box15, GK19] that the Ekedahl–Oort stratification extends to a stratification of the minimal compactification of $Y_{K^p, \mathrm{Iw} K_p, 1}$, and that each (open) stratum is affine. Moreover the superspecial locus does not intersect the boundary. It follows that any component of any Ekedahl–Oort stratum contains a superspecial point in its closure. By the explicit description of the Ekedahl–Oort stratification recalled above, the p-rank strata in the statement of the lemma are also Ekedahl–Oort strata. □

Now we will switch to Iwahori level and consider the Kottwitz–Rapoport stratification, see for instance [NG02]. Let $K_p^{\mathrm{Iw}} = \prod_{v|p} \mathrm{Iw}(v)$. Then $Y_{K^p, \mathrm{Iw} K_p^{\mathrm{Iw}}, 1}$ and its local model $M^{\mathrm{loc}}_{K_p^{\mathrm{Iw}}, 1} = \prod_{v|p} M^{\mathrm{loc}}_{\mathrm{Iw}(v), 1}$ carry a Kottwitz–Rapoport stratification. In fact, there is a Kottwitz–Rapoport stratification of $M^{\mathrm{loc}}_{\mathrm{Iw}(v), 1}$ and the stratification of $M^{\mathrm{loc}}_{K_p^{\mathrm{Iw}}, 1}$ is simply the product stratification. The strata of $M^{\mathrm{loc}}_{\mathrm{Iw}(v)}$ are indexed by a set $\mathrm{Adm}(\mu)$ of cardinality 13. This set, as well as the partial ordering given by closure, is pictured in [Yu08, p. 1273].

We will use below the following argument, which is a consequence of the theory of local models. If C is an irreducible component of the Kottwitz–Rapoport stratum labeled by $w \in \mathrm{Adm}(\mu)^{S_p}$, then the closure $\overline{C}$ has a decomposition into strata:

$$\overline{C} = \coprod_{w' \leq w} \overline{C}_{w'}.$$

A priori the strata $\overline{C}_{w'}$ might be empty, although it is expected that they are always nonempty. However, the theory of local models implies that if $\overline{C}_{w'}$ is nonempty, then so is $\overline{C}_{w''}$ for any w'' satisfying $w' \leq w'' \leq w$.

We will not need to recall in detail the definition of the Kottwitz–Rapoport stratification. We do recall that, as explained in [Yu08], the Kottwitz–Rapoport invariant determines whether the groups of order p, H_w and L_w/H_w, are étale, multiplicative, or connected-connected (and so in particular the Kottwitz–Rapoport invariant determines the p-rank of $\mathcal{G}_w$, a theorem of Genestier–Ngô). Conversely these invariants determine the Kottwitz–Rapoport invariant when the p-rank of $\mathcal{G}_w$ is not 0. All of this is recorded in the table in [Yu08, p. 1276].

We will use the following points:

- There is a Kottwitz–Rapoport condition, $s_2 s_1 s_2 \tau$ in [Yu08], which corresponds to the condition that $\mathcal{G}_w$ is ordinary and $L_w = \mathcal{G}[F]$ (equivalently L_w is multiplicative).
- There is a Kottwitz–Rapoport condition, $s_1 s_2 \tau$ in [Yu08], which corresponds to the condition that $\mathcal{G}_w$ has p-rank 1, H_w is multiplicative, and $L_w = \mathcal{G}[F]$ (equivalently H_w is multiplicative and L_w/H_w is connected-connected).
- There are three Kottwitz–Rapoport conditions, $\tau, s_1\tau$, and $s_2\tau$ in [Yu08], which have p-rank 0 and are in the closure of the first stratum recalled above. We observe crucially that they are also all in the closure of the second stratum recalled above. We refer to these three strata as the canonical p-rank 0 Kottwitz–Rapoport strata (here "canonical" refers to the fact that $L_w = \mathcal{G}_w[F]$ is the canonical subgroup of $\mathcal{G}_w$).

For $J \subseteq S_p$ we write $Y^{=_{J^c}2,=_J 1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ for the locus in $Y_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ where for $w \in J^c$, $\mathcal{G}_w$ is ordinary and $L_w = \mathcal{G}_w[F]$, while for $w \in J$, $\mathcal{G}_w$ has p-rank 1, H_w is multiplicative, and $L_w = \mathcal{G}_w[F]$. By what we have just recalled, this is a Kottwitz–Rapoport stratum.

Lemma 8.2.11. *For $J \subseteq J' \subseteq S_p$, any irreducible component of $Y^{=_{J^c}2,=_J 1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ contains a point of $Y^{=_{(J')^c}2,=_{J'} 1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ in its closure.*

Proof. Let $\pi : Y_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1} \to Y_{K^{p,\mathrm{Iw}}K_p,1}$ be the projection from Iwahori to spherical level. It is proper, and the Kottwitz–Rapoport stratum $Y^{=_{J^c}2,=_J 1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ maps finitely onto the p-rank stratum $Y^{=_{J^c}2,=_J 1}_{K^{p,\mathrm{Iw}}K_p,1}$ (the fibres correspond to the $p+1$ choices of H_w for $w \in J^c$). If C is an irreducible component of $Y^{=_{J^c}2,=_J 1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$, then $\pi(C)$ is an irreducible component of $Y^{=_{J^c}2,=_J 1}_{K^{p,\mathrm{Iw}}K_p,1}$. By Lemma 8.2.10, the closure $\overline{\pi(C)}$ contains a point which is p-rank 0 for all $w \in S_p$. By the properness of π it follows that the closure $\overline{C}$ contains a point which is p-rank 0 for all $w \in S_p$.

By what we have shown, in the closure $\overline{C}$, at least one of the canonical p-rank 0 Kottwitz–Rapoport strata is nonempty. Now we apply the argument with local models and the explicit description of the closure relations between the strata recalled above to conclude. □

Remark 8.2.12. One could give a more direct proof of Lemma 8.2.11, avoiding the consideration of the Ekedahl–Oort stratification and the superspecial locus at spherical level, if one knew that the Kottwitz–Rapoport stratification of $Y_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}}}$

extended to a stratification of the minimal compactification, for which the (open) strata are quasi-affine. However we lack a reference for these facts.

Proof of Lemma 8.2.8. Let $\pi : Y_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1} \to Y_{K^{p,\mathrm{Iw}}K_p(I),1}$ be the projection from Iwahori to $K_p(I)$ level. On $Y_{K^{p,\mathrm{Iw}}K_p(I),1}$, π has a "canonical section" $s : Y_{K^{p,\mathrm{Iw}}K_p(I),1} \to Y_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$, defined by taking $L_w = \mathcal{G}_w[F]$ for $w \in I$ (recall that on $Y_{K^{p,\mathrm{Iw}}K_p(I),1}$, H_w is multiplicative by definition, and hence $H_w \subseteq \mathcal{G}_w[F]$). It follows that for $J \subseteq I$, s and π define mutually inverse isomorphisms between $Y^{=_{J^c}2,=_J1,m-can}_{K^{p,\mathrm{Iw}}K_p^{\mathrm{Iw}},1}$ and $Y^{I,=_{J^c}2,=_J1}_{K^{p,\mathrm{Iw}}K_p(I),1}$. We deduce the following statement from Lemma 8.2.11: for $J \subseteq J' \subseteq I$, every irreducible component of $Y^{I,=_{J^c}2,=_J1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ contains a point of $Y^{I,=_{(J')^c}2,=_{J'}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ in its closure.

Applying this with $J = \{v_2\}$, $J' = \{v_1, v_2\}$ and $J = \emptyset$, $J' = \{v_1\}$ we conclude that any irreducible component of $Y^{=_{S_p\setminus\{v_2\}}2,=_{v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ contains a point of $Y^{I,=_{v_1,v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ in its closure, and any irreducible component of $Y^{=_{S_p}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$ contains a point of $Y^{I,=_{v_1}1,=_{v_2}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$ in its closure. Finally as recalled at the start of section 4.1, $Y^{=_{S_p\setminus\{v_2\}}2,=_{v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ is dense in $Y^{=_{v_2}1}_{K^{p,\mathrm{Iw}}K_p(I),1}$ and $Y^{=_{S_p}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$ is dense in $Y^{=_{v_2}2}_{K^{p,\mathrm{Iw}}K_p(I),1}$, and the lemma follows. □

8.3. **Solvable base change.** We will use solvable base change to deduce our main modularity lifting theorem from Corollary 8.2.2. We firstly prove a couple of preparatory lemmas, beginning with the following well-known result.

Lemma 8.3.1. *Let K be a number field, and let $\rho : G_K \to \mathrm{GL}_4(\overline{\mathbf{Q}}_p)$ be an irreducible representation which preserves a generalized symplectic form with similitude character ν. Then either ν is uniquely determined by ρ, or, if ρ also admits a similitude character $\nu\psi$ with $\psi \neq 1$, then ψ has finite order and ρ is reducible over a quadratic subfield of the fixed field of ψ and hence also over the fixed field of ψ.*

Proof. Let V denote the underlying representation of ρ, and let ν and $\nu\psi$ denote two possible similitude characters. Then there is an inclusion $\nu \oplus \nu\psi \subset \mathrm{Hom}(V^*, V)$, or equivalently, $1 \oplus \psi \subset \mathrm{Hom}(V^*(\nu), V)$. It follows that $V \simeq V^*(\nu)$ and $V \simeq V^*(\nu\psi)$, and thus $V \simeq V(\psi)$, and also $V \simeq V(\psi) \simeq V(\psi^2)$. By comparing determinants, it follows that ψ^4 is trivial, and hence either ψ or ψ^2 is a quadratic character η such that $V \simeq V(\eta)$ and hence $1 \oplus \eta \subset \mathrm{Hom}(V, V)$. By Schur's Lemma, V becomes reducible over the fixed field of η, which by construction is a quadratic subfield of the fixed field of ψ. □

We now prove a slightly technical lemma on solvable base change; it is an analogue of [BLGHT11, Lem. 1.3] for GSp_4, but the proof is slightly more involved.

Lemma 8.3.2. *Suppose that $p > 2$ splits completely in the totally real field $F/\mathbf{Q}$. Let F'/F be a solvable extension of totally real fields. Suppose that $\rho : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ satisfies:*

(1) $\nu \circ \rho = \varepsilon^{-1}$.
(2) For all $v|p$, $\rho|_{G_{F_v}}$ is p-distinguished weight 2 ordinary.
(3) The representation $\overline{\rho}$ is vast and tidy.
(4) $\rho|_{G_{F'}}$ is irreducible. Furthermore, there is an ordinary automorphic representation π' of $\mathrm{GSp}_4(\mathbf{A}_{F'})$ of parallel weight 2 and central character $|\cdot|^2$,

such for every finite place w of F' we have

$$\mathrm{WD}(\rho|_{G_{F'_w}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_w \otimes |\nu|^{-3/2}).$$

(So in particular, $\rho_{\pi',p} \cong \rho|_{G_{F'}}$.)

Then ρ is modular. More precisely, there is an ordinary automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 and central character $|\cdot|^2$, with $\rho_{\pi,p} \cong \rho$. Furthermore, for every finite place v of F we have

$$\mathrm{WD}(\rho|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2}).$$

Proof. Since $\rho_{\pi',p}$ is irreducible, π' must be of general type in the sense of [Art04], so that it corresponds to a cuspidal automorphic representation Π' of $\mathrm{GL}_4(\mathbf{A}_{F'})$. By induction we may reduce to the case that F'/F is cyclic of prime degree, in which case it follows from [AC89, Thm. 4.2 of §3] that there is an automorphic representation Π of $\mathrm{GL}_4(\mathbf{A}_F)$ with $\mathrm{BC}_{F'/F}(\Pi) = \Pi'$.

We can write

$$L^S(s, \Pi', \bigwedge^2 \otimes |\cdot|^{-2}) = \prod_{\psi} L^S(s, \Pi, \bigwedge^2 \otimes |\cdot|^{-2}\psi^{-1})$$

where the product is over the characters ψ of $\mathbf{A}_F^\times / F^\times N_{F'/F}\mathbf{A}_{F'}^\times$. The left hand side has a simple pole at $s = 1$ (by the assumption that Π' is the transfer of π'), while by the main result of [Sha97], all but at most one factor on the right hand side is holomorphic and non-vanishing at $s = 1$. Thus some factor on the right hand side must also have a simple pole at $s = 1$, say $L^S(s, \Pi, \bigwedge^2 \otimes |\cdot|^{-2}\psi^{-1})$.

It follows from Theorem 2.9.3 that Π is the transfer of a cuspidal automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_F)$ with central character $|\cdot|^2\psi$. Since $\mathrm{BC}_{F'/F}(\Pi) = \Pi'$, we see that π is of parallel weight 2. Letting $\rho_{\pi,p} : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ be the Galois representation corresponding to π (whose existence follows from [Mok14, Thm. 3.5] exactly as in the proof of Theorem 2.7.2), we have $\rho_{\pi,p}|_{G_{F'}} \cong \rho|_{G_{F'}}$, so that (since F'/F is cyclic of prime degree, and $\rho|_{G_{F'}}$ is irreducible) $\rho_{\pi,p}$ differs from ρ by a twist by a character of $\mathrm{Gal}(F'/F)$.

Replacing π by the corresponding twist, we may assume that $\rho_{\pi,p}$ and ρ are isomorphic when considered as representations valued in $\mathrm{GL}_4(\overline{\mathbf{Q}}_p)$. We claim that we necessarily have $\nu \circ \rho = \nu \circ \rho_{\pi,p} = \varepsilon^{-1}$, so that π has central character $|\cdot|^2$. Indeed, this follows from Lemma 8.3.1, since it holds after restriction to $G_{F'}$, and $\rho|_{G_{F'}}$ is irreducible by assumption. So $\rho_{\pi,p} \cong \rho$, as required.

Since we have assumed that $\overline{\rho}$ is vast and tidy, it follows from Corollary 7.9.6 that for every finite place v of F we have

$$\mathrm{WD}(\rho|_{G_{F_v}})^{\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})^{\mathrm{ss}}.$$

It remains to check that the monodromy operators agree; but this may be checked after base change, and since π' is the base change of π, it follows from the assumption that $\mathrm{WD}(\rho|_{G_{F'_w}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_w \otimes |\nu|^{-3/2})$. □

8.4. **The main modularity lifting theorem.** We now prove our main modularity lifting theorem.

Theorem 8.4.1. *Suppose that $p \geq 3$ splits completely in the totally real field $F/\mathbf{Q}$. Suppose that $\rho : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ satisfies:*

(1) $\nu \circ \rho = \varepsilon^{-1}$.

(2) *The representation $\overline{\rho}$ is vast and tidy in the sense of Definitions 7.5.6 and 7.5.11.*
(3) *For all $v|p$, $\rho|_{G_{F_v}}$ is p-distinguished weight 2 ordinary in the sense of Definition 7.3.1.*
(4) *There exists π of parallel weight 2 and central character $|\cdot|^2$, which is ordinary at all $v|p$, such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}$.*
(5) *For all finite places v of F, $\rho|_{G_{F_v}}$ and $\rho_{\pi,p}|_{G_{F_v}}$ are pure.*

Then ρ is modular. More precisely, there is an ordinary automorphic representation π' of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 and central character $|\cdot|^2$ which satisfies $\rho_{\pi',p} \cong \rho$. Furthermore, for every finite place v of F we have

$$\mathrm{WD}(\rho|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_v \otimes |\nu|^{-3/2}).$$

Proof. Choose a solvable extension of totally real fields F'/F, linearly disjoint from $F^{\ker \overline{\rho}}$ over F, with the following properties:

- p splits completely in F'.
- At every place w of F' lying over a place $v \nmid p$ of F for which π_v or $\rho|_{G_{F_v}}$ is ramified, $\overline{\rho}|_{G_{F'_w}}$ is trivial, $\rho|_{G_{F'_w}}$ has only unipotent ramification, and $q_w \equiv 1 \pmod{p^2}$.
- There is an automorphic representation π' of $\mathrm{GSp}_4(\mathbf{A}_{F'})$ of parallel weight 2 which is a base change of π (in the sense that for each finite place w of F', lying over a place v of F, we have $\mathrm{rec}_{\mathrm{GT},p}(\pi') = \mathrm{rec}_{\mathrm{GT},p}(\pi)|_{W_{F'_w}}$). Furthermore, for all finite places w of F' we have $(\pi'_w)^{\mathrm{Iw}(w)} \neq 0$.

(The last property can be arranged by [Mok14, Prop. 4.13].) Then $\rho|_{G_{F'}}$ satisfies Hypothesis 7.13.1, so the result follows from Corollary 8.2.2 (applied to $\rho|_{G_{F'}}$) and Lemma 8.3.2. □

8.5. **Base change and automorphy lifting.** Throughout the paper, we have fixed the similitude factor of our Galois representations to be ε^{-1}, in order to streamline both the notation and some arguments. We now explain how to use base change to relax this condition in our main automorphy lifting theorem. We do not use this result elsewhere in the paper, so we have contented ourselves with a slightly ugly statement, and with a sketch of the proof.

Definition 8.5.1. We say that a representation $\rho : G_{\mathbf{Q}_p} \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ is *twisted p-distinguished weight* 2 *ordinary* if it is an unramified twist of a representation which is p-distinguished weight 2 ordinary in the sense of Definition 7.3.1. Similarly, we say that an admissible representation π_p of $\mathrm{GSp}_4(\mathbf{Q}_p)$ is *twisted ordinary* if it is an unramified twist of an ordinary representation.

Theorem 8.5.2. *Suppose that $p \geq 3$ splits completely in the totally real field $F/\mathbf{Q}$. Suppose that $\rho : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ satisfies:*

(1) *$\nu \circ \rho = \chi\varepsilon^{-1}$, where χ is a totally even finite order character, which is unramified at all places dividing p.*
(2) *The representation $\overline{\rho}$ is vast and tidy in the sense of Definitions 7.5.6 and 7.5.11.*
(3) *For all $v|p$, $\rho|_{G_{F_v}}$ is twisted p-distinguished weight 2 ordinary.*
(4) *There exists π of parallel weight 2, which is twisted ordinary at all $v|p$, such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}$.*
(5) *For all finite places v of F, $\rho|_{G_{F_v}}$ and $\rho_{\pi,p}|_{G_{F_v}}$ are pure.*

Then ρ is modular. More precisely, there is a twisted ordinary automorphic representation π' of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 which satisfies $\rho_{\pi',p} \cong \rho$. Furthermore, for every finite place v of F we have

$$\mathrm{WD}(\rho|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi'_v \otimes |\nu|^{-3/2}).$$

Proof. Let χ' be the finite order character $G_F \to \overline{\mathbf{Q}}_p^\times$ such that $\nu \circ \rho_\pi = \chi'\varepsilon^{-1}$. Note that χ' is totally even (since we have $\overline{\chi}' = \overline{\chi}$ by assumption). We can choose a quadratic extension of totally real fields F'/F, linearly disjoint from $F^{\ker \overline{\rho}}$ over F, such that:

- p splits completely in F', and
- there are finite order characters $\psi, \psi' : G_{F'} \to \overline{\mathbf{Q}}_p^\times$ such that $\chi|_{G_{F'}} = \psi^2$, $\chi'|_{G_{F'}} = (\psi')^2$.

Indeed, the obstruction to taking the square root of a character is in the 2-torsion of the Brauer group, and there are no obstructions to taking a square root of either χ or χ' at the places dividing p (because both characters are unramified at such places) or at the infinite places (because χ, χ' are totally even).

Let $\pi_{F'}$ be the base change of π to F'. Since $\rho|_{G_{F'}} \otimes \psi^{-1}$, $\pi_{F'} \otimes (\psi')^{-1} \circ \mathrm{Art}_{F'}$ satisfy the hypotheses of Theorem 8.4.1, it follows that $\rho|_{G_{F'}} \otimes \psi$ is modular, so $\rho|_{G'_F}$ itself is modular. The result follows from Lemma 8.3.2 (or rather, from an obvious generalization of this lemma to the case of more general central characters, which may be proved in the same way). □

9. Potential modularity of abelian surfaces

We now use the potential automorphy methods introduced in [Tay02] to prove the potential modularity of abelian surfaces. It is presumably possible to follow [Tay02, §1] quite closely, but we instead make use of potential modularity results for GL_2 and the local to global principle of [Cal12, §3] (see also [MB90, Thm. 1.2]).

9.1. Compatible systems and potential automorphy.

Recall that the notion of a C-algebraic automorphic representation is defined in [BG14], and in the case of automorphic representations of GL_n, this definition agrees with the notion of an algebraic automorphic representation defined in [Clo90].

Definition 9.1.1. Let K be a number field and let $\mathcal{R}$ be a strictly compatible system of representations of G_K. We say that $\mathcal{R}$ is *automorphic* if there is an automorphic representation Π of $\mathrm{GL}_n(\mathbf{A}_K)$, with the properties that:

(1) Π is an isobaric direct sum of cuspidal automorphic representations $\boxplus_{i=1}^r \Pi_i$ where each Π_i is a C-algebraic cuspidal automorphic representation of some $\mathrm{GL}_{n_i}(\mathbf{A}_K)$.
(2) The fixed field M_Π of the subgroup of $\mathrm{Aut}(\mathbf{C})$ consisting of those $\sigma \in \mathrm{Aut}(\mathbf{C})$ with ${}^\sigma\Pi^\infty \cong \Pi^\infty$ is a number field.
(3) For each finite place v of K, $\mathrm{WD}_v(\mathcal{R}) = \oplus_i \mathrm{rec}(\Pi_{i,v}|\det|_v^{(1-n_i)/2})$.

Remark 9.1.2. There are many (conjecturally equivalent) variants of Definition 9.1.1 that could be made. The definition is in some sense redundant, because condition (2) is implied by condition (3); indeed, by the definition of a compatible system, it follows that for all but finitely many v, Π_v is an unramified principal series representation, defined over a number field which may be chosen independently of v. Condition (2) then follows from strong multiplicity one for isobaric

representations [JS81]. The reason that we have chosen to include the condition separately is that conjecturally (see [Clo90]) condition (1) implies condition (2), and also implies the existence of a compatible system $\mathcal{R}$ satisfying condition (3).

In fact, the only cases of Definition 9.1.1 that we will need to consider are those where either:

(1) Each Π_i is regular algebraic, or
(2) K is totally real, Π is cuspidal, and Π is the transfer to GL_4 of a cuspidal automorphic representation of GSp_4 of parallel weight 2 and central character $|\cdot|^2$.

In either case, condition (2) is satisfied by [Clo90, Thm. 3.13] and [BHR94, Thm. 3.2.2]) respectively.

Remark 9.1.3. The reader may wonder why we did not demand an analogue of condition (3) of Definition 9.1.1 at the infinite places. One reason is that we do not need to do so, as condition (3) already determined Π uniquely (indeed, as in Remark 9.1.2, this is already true if one only considers condition (3) at all but finitely many places). The main reason that we do not make a requirement at the infinite places is that (in keeping with the literature) our definition of a compatible system does not include a requirement that the l-adic representations are compatible on complex conjugations, which makes it harder to formulate a precise compatibility. One could certainly ask (as in [BG14]) that the Hodge–Tate weights of $\mathcal{R}$ correspond to the infinitesimal character of Π, but to save introducing additional notation and terminology we have not done so.

Remark 9.1.4. As explained in Remark 9.1.2, condition (3) of Definition 9.1.1 at all but finitely many places v determines Π uniquely. One might ask whether if this condition holds for all but finitely many v, it necessarily holds for all v. In general this is a hard problem; indeed even if (3) is known up to semisimplification, it is often difficult to show that the monodromy operators agree. If however $\mathcal{R}$ is pure and Π is generic then the agreement of monodromy operators is automatic; we will use this fact in our arguments below.

If $\mathcal{R}$ and Π are as in Definition 9.1.1, we as usual have Gamma factors $L_v(\Pi, s)$ for each place $v|\infty$ of K, and we set

$$\Lambda_\Pi(\mathcal{R}, s) = L(\mathcal{R}, s) \prod_{v|\infty} L_v(\Pi, s).$$

This is of course just the usual completed L-function of Π, but we have included $\mathcal{R}$ in the notation to emphasize that the L-functions of Π and $\mathcal{R}$ agree – note that here it is important that we know Definition 9.1.1 (3) at all finite places, and not just at almost all places, or up to semisimplification. As noted above, since we do not *a priori* demand any local-global compatibility at ∞ for our compatible system $\mathcal{R}$, we use the automorphic representation Π in this definition mostly as a convenient way to write down the correct Gamma factors at infinity. For those who find this notation unpleasant, note that — in the restrictive context of abelian surfaces over totally real fields — we defined a function $\Lambda(\mathcal{R}, s)$ in (2.8.4) by explicitly writing down the Gamma factors in question, and then (for all the A and Π that arise in this paper) we indeed have equalities $\Lambda(\Pi, s) = \Lambda_\Pi(\mathcal{R}_A, s) = \Lambda(\mathcal{R}_A, s)$.

We also have an epsilon factor $\varepsilon(\Pi)$, and a conductor $N(\Pi)$, and by [GJ72, Cor. 13.8], $\Lambda_\Pi(\mathcal{R}, s)$ admits a meromorphic continuation to the entire complex plane,

and satisfies the functional equation

$$\Lambda_\Pi(\mathcal{R}, s) = \varepsilon(\Pi) N(\Pi)^{-s} \Lambda_\Pi(\mathcal{R}^\vee, 1-s). \tag{9.1.5}$$

Definition 9.1.6. Let A/K be an abelian variety. We say that A is *automorphic* if $\mathcal{R}_A$ is automorphic in the sense of Definition 9.1.1. We say that it is *potentially automorphic* if there is a finite extension of number fields L/K such that $\mathcal{R}_A|_{G_L}$ is automorphic.

Remark 9.1.7. If A/K is an abelian variety, and A is automorphic with the corresponding Π being of the form considered in Remark 9.1.2, then the Gamma factors $L_v(\Pi, s)$ and the conductor $N(\Pi)$ agree with those defined for the compatible system $\mathcal{R}_A$ in §2.8. Indeed, for the Gamma factors this is a direct consequence of the definitions, and the conductor respects the local Langlands correspondence. In particular, we have $\Lambda_\Pi(\mathcal{R}_A, s) = \Lambda(\mathcal{R}_A, s)$, and the functional equations (2.8.5) and (9.1.5) agree; so $\Lambda(\mathcal{R}_A, s)$ satisfies the expected meromorphic continuation and functional equation.

Definition 9.1.8. Let F be a totally real. We say that a representation $\rho : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ is *modular* if there is a cuspidal automorphic representation of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 and central character $|\cdot|^2$ for each finite place v of F we have $\mathrm{WD}(\rho_{\pi,p}|_{G_{F_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}_{\mathrm{GT},p}(\pi_v \otimes |\nu|^{-3/2})$; in particular, $\rho \cong \rho_{\pi,p}$.

We say that r is *potentially modular* if there is a finite Galois extension F'/F of totally real fields such that $r|_{G_{F'}}$ is modular.

If A/F is an abelian surface, we say that A is modular (resp. potentially modular) if $\rho_{A,p}$ is modular (resp. potentially modular) for some (equivalently, for any) prime p.

Remark 9.1.9. The relationship between the definitions of what it means for an abelian surface A/F to be (potentially) automorphic or modular is somewhat complicated, because of the various possibilities in Arthur's classification of the discrete spectrum of $\mathrm{GSp}_4(\mathbf{A}_F)$. In this paper we will only show that A is (potentially) modular if the corresponding automorphic representation of GSp_4 is of general type, in which case A is also (potentially) automorphic, essentially by the definition of "general type"; note that this is the case considered in Remark 9.1.2 (2).

We now prove some technical lemmas that we will use in proving our main potential automorphy/modularity results. A weakly compatible system $\mathcal{R}$ is defined to be *irreducible* if there is a set $\mathcal{L}$ of rational primes of Dirichlet density 1 such that for $\lambda | l \in \mathcal{L}$ the representation r_λ is irreducible. We say that it is *strongly irreducible* if for all finite extensions F'/F the compatible system $\mathcal{R}|_{G_{F'}}$ is irreducible. If $n = 2$, then we say that $\mathcal{R}$ has weight 0 if $H_\tau(\mathcal{R}) = \{0, 1\}$ for each τ, and we say that $\mathcal{R}$ is odd if $\det \mathcal{R}(c_v) = -1$ for all $v|\infty$. If $\boldsymbol{\pi}$ is a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_F)$ of weight 0, then $\mathcal{R}(\boldsymbol{\pi})$ is odd and has weight 0.

We have the following standard lemma.

Lemma 9.1.10. *Let $\mathcal{R}$ be a rank two weakly compatible system.*

(1) The following are equivalent:
  (a) $\mathcal{R}$ is irreducible.
  (b) For all λ the representation r_λ is irreducible.
  (c) For some λ the representation r_λ is irreducible.

(2) If $\mathcal{R}$ is irreducible and regular, then the following are equivalent:

(a) $\mathcal{R}$ *is strongly irreducible.*
(b) $\mathrm{Sym}^2 \mathcal{R}$ *is irreducible.*
(c) For all λ, $\mathrm{Sym}^2 r_\lambda$ *is irreducible.*
(d) For some λ, $\mathrm{Sym}^2 r_\lambda$ *is irreducible.*

If these equivalent conditions do not hold, then there is a quadratic extension F'/F *and a weakly (equivalently, strongly) compatible system* $\mathcal{X}$ *of characters of* $G_{F'}$ *such that*

$$\mathcal{R} \cong \mathrm{Ind}_{G_{F'}}^{G_F} \mathcal{X}.$$

(3) If $\mathcal{R}$ *is strongly irreducible and regular, then for a density one set of primes* l *of* $\mathbf{Q}$*, if* $\lambda | l$ *is a place of* M*, then the image of* $\overline{r}_\lambda$ *contains* $\mathrm{SL}_2(\mathcal{O}_M/\lambda)$.

Proof. This is well known. Part (1) is [ACC+18, Lem. 7.1.1], and part (3) is [ACC+18, Lem. 7.1.3]. For part (2), note that by [ACC+18, Lem. 7.1.2], either $\mathcal{R}$ is strongly irreducible, or we can write $\mathcal{R} \cong \mathrm{Ind}_{G_{F'}}^{G_F} \mathcal{X}$. It follows that if $\mathcal{R}$ is not strongly irreducible, then $\mathrm{Sym}^2 r_\lambda$ is reducible for every λ.

Conversely, if $\mathrm{Sym}^2 r_\lambda$ is reducible for some λ, then there is a nontrivial character ψ such that $r_\lambda \cong r_\lambda \otimes \psi$. Considering determinants, ψ is a quadratic character. Letting F'/F be the quadratic extension corresponding to ψ, it follows from Schur's lemma that $r_\lambda|_{G_{F'}}$ is reducible, so by part (1), $\mathcal{R}$ is not strongly irreducible. □

We now use a standard trick with restriction of scalars to give some slight improvements to some applications of the theorem of Moret-Bailly.

Proposition 9.1.11. *Let* F_1/F *be a finite extension of totally real fields, and let* $p, q > 2$ *be distinct primes which split completely in* F_1*. Let* $\overline{r} : G_{F_1} \to \mathrm{GL}_2(\overline{\mathbf{F}}_q)$ *be a representation with determinant* $\overline{\varepsilon}^{-1}$.

Suppose that for each place $v|q$ *of* F_1*,* $\overline{r}|_{G_{F_{1,v}}}$ *is of the form* $\begin{pmatrix} \lambda_{\overline{\alpha}_v} & 0 \\ 0 & \varepsilon^{-1}\lambda_{\overline{\alpha}_v}^{-1} \end{pmatrix}$. *Suppose also that* $\overline{r}$ *is unramified at all places above* p.

Let $F^{(\mathrm{avoid})}/F$ *be a finite extension. Then there is a finite Galois extension* F'/F *of totally real fields in which* p *and* q *split completely and which is linearly disjoint from* $F_1F^{(\mathrm{avoid})}/F$*, and a* q*-ordinary cuspidal automorphic representation* $\boldsymbol{\pi}$ *of* $\mathrm{GL}_2(\mathbf{A}_{F_1F'})$ *of weight* 0 *and trivial central character which is unramified at all places dividing* pq *and which satisfies* $\overline{\rho}_{\boldsymbol{\pi},q} \cong \overline{r}|_{G_{F_1F'}}$.

Proof. In the case $F_1 = F$, this is a straightforward consequence of [Sno09, Thm. 8.2.1]. Indeed, in Snowden's notation, we take $\overline{\rho} = \overline{r}^\vee$, $\psi = 1$, we let S consist of the places dividing pq and we let t assign the type A at places lying over q and type AB at places lying over p. To prove the general case, one simply replaces the scheme X to which Snowden applies the theorem of Moret-Bailly with the restriction of scalars $\mathrm{Res}_{F_1/F} X_{F_1}$. □

Proposition 9.1.12. *Let* G *be a finite group, let* $E/\mathbf{Q}$ *be a finite extension, and let* S *be a finite set of places of* E*. Let* E'/E *be a finite extension, and let* $F^{(\mathrm{avoid})}/E$ *be a finite extension, linearly disjoint from* E'/E.

Let S'/S *be the set of places of* E' *lying over places of* S*. For each finite place* $v \in S'$*, let* H'_v/E'_v *be a finite Galois extension together with a fixed inclusion* $\phi_v : \mathrm{Gal}(H'_v/E'_v) \hookrightarrow G$ *with image* D_v*. For each real infinite place* $v \in S'$*, let* $c_v \in G$ *be an element of order dividing* 2.

Then there exists a number field K/E and a finite Galois extension of number fields L/K such that if we set $K' = KE'$, $L' = LE'$, then

1. *There is an isomorphism* $\mathrm{Gal}(L'/K') = G$.
2. *L'/E is linearly disjoint from $E'F^{(\mathrm{avoid})}/E$.*
3. *All places in S' split completely in K'.*
4. *For all finite places w of K' above $v \in S'$, the local extension L'_w/K'_w is equal to H'_v/E'_v. Moreover, there is a commutative diagram:*

$$\begin{array}{ccc} \mathrm{Gal}(L'_w/K'_v) & \longrightarrow & D_w \subset G \\ \| & & \| \\ \mathrm{Gal}(H'_v/E'_v) & \xrightarrow{\phi_v} & D_v \subset G \end{array}$$

5. *For all real places $w|\infty$ of K' above $v \in S'$, complex conjugation $c_w \in G$ is conjugate to c_v.*

Proof. The case $E' = E$ is [Cal12, Prop. 3.2] (see also [MB90, Thm. 1.2]). The general case may be proved in exactly the same way, by replacing the application of [Cal12, Thm. 3.1] to (an open subscheme of) X_G/E with an application of it to $\mathrm{Res}_{E'/E} X_G$. □

9.2. **Abelian surfaces.** We begin by recalling some results from [FKRS12] and [Joh17], which allow us to deal with various cases where the abelian surfaces have extra endomorphisms, and can be handled with the potential automorphy theorems of [BLGGT14b]. Let A/F be an abelian surface over a totally real field F, and let L/F be the minimal extension over which all its endomorphisms are defined. This is a Galois extension, and following [FKRS12] we say that the *Galois type* of A/F is the $\mathrm{Gal}(L/F)$-module $\mathrm{End}_L(A) \otimes_{\mathbf{Z}} \mathbf{R}$. These possible Galois types are classified in [FKRS12], and they are divided up into 6 families **A**-**F**.

The precise classification of monodromy groups in these references is not actually strictly necessary for our purposes. Write $\{\rho_{A,l}\}$ for the compatible system of Galois representations $\{H^1(A_{\overline{F}}, \mathbf{Q}_l)\}$. In practice, it suffices to know that the l-adic representations $\rho_{A,l}$ fall into precisely one of the following categories independently of l:

1. strongly irreducible (type **A**),
2. reducible (type $\mathbf{B}[C_1]$, **C**, $\mathbf{E}[C_n]$, some **D**, some **F**),
3. potentially abelian but not reducible (of type the remaining **D** and **F** cases),
4. induced from a quadratic extension K/F but not potentially abelian, in which case either:
   (a) the two 2-dimensional representations over K are equivalent up to twist (type $\mathbf{E}[D_n]$), or
   (b) the two 2-dimensional representations over K are not equivalent up to twist (type $\mathbf{B}[C_2]$).

Proposition 9.2.1. *Suppose that A/F is not of type* **A** *or* $\mathbf{B}[C_2]$. *Then A is potentially automorphic.*

Proof. We freely use the discussion of [Joh17, §4]. In cases **D**, **F**, A_L is of CM type, so the compatible system $\mathcal{R}_A$ is potentially abelian, and in particular potentially automorphic.

In cases $\mathbf{B}[C_1]$, $\mathbf{C}$, and the cases of type $\mathbf{E}$ other than those of type $\mathbf{E}[D_n]$, it follows from the discussions at the beginnings of [Joh17, §4.2, 4.4, 4.5] that we can write $\mathcal{R}_A = \mathcal{R}_A^1 \oplus \mathcal{R}_A^2$ where each $\mathcal{R}_A^i$ is an irreducible, odd, weight 0 weakly compatible system of rank 2 l-adic representations of G_F. The potential automorphy of $\mathcal{R}_A$ therefore follows from [BLGGT14b, Thm. 5.4.1].

It remains to treat the case that A is of type $\mathbf{E}[D_n]$. In this case, as explained in [Joh17, §4.5-4.6], there is a quadratic extension F'/F and a strongly irreducible weakly compatible system $\mathcal{S} = \{s_l\}$ of weight 0 representations of $G_{F'}$ which is defined over $\mathbf{Q}$ such that $\mathcal{R}_A \cong \operatorname{Ind}_{G_{F'}}^{G_F} \mathcal{S}$. Furthermore, there is a finite order character δ of $G_{F'}$ such that if we write $\operatorname{Gal}(F'/F) = \{1, \sigma\}$, then $s_l^\sigma \cong s_l \otimes \delta_l$.

It follows from Lemma 9.1.10 (2) and [BLGGT14b, Prop. 5.3.2] that for a density one set of primes l, $\operatorname{Sym}^2 \overline{s}_l|_{G_{F'(\zeta_l)}}$ is irreducible, l is unramified in F', and both s_l and ε_l are crystalline at all primes above l. Fix one such $l > 7$.

Since $\operatorname{Proj} s_l^\sigma \cong \operatorname{Proj} s_l$, it follows from Schur's lemma that $\operatorname{Proj} s_l$ extends to a representation $G_F \to \mathrm{PGL}_2(\overline{\mathbf{Q}}_l)$. By [Pat19, Lem. 2.3.17, 2.7.4], we may lift this to a representation $\tilde{r} : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_l)$ which is unramified at all but finitely many places, and is Hodge–Tate at all places dividing l, with Hodge–Tate weights $(0, 1)$. By construction, there is a character $\psi : G_{F'} \to \overline{\mathbf{Q}}_l^\times$ such that $\tilde{r}|_{G_{F'}} = s_l \otimes \psi$. Since ψ is Hodge–Tate of weight 0, it has finite order.

Since l is unramified in F' and s_l is crystalline at all primes above l, after possibly replacing $\tilde{r}$ by a twist by a finite order character, we may assume that it is crystalline at all places dividing l. By [CG13, Prop. 2.5], $\tilde{r}$ is odd, so by [BLGGT14b, Thm. 4.5.1], $\tilde{r}$ is potentially automorphic. Since

$$\rho_{A,l} \cong \operatorname{Ind}_{G_{F'}}^{G_F} s_l \cong \operatorname{Ind}_{G_{F'}}^{G_F} \left(\tilde{r}|_{G_{F'}} \otimes \psi^{-1}\right),$$

it follows that $\mathcal{R}_A$ is potentially automorphic, as required. □

We say that A/F is *challenging* if it has type $\mathbf{A}$ (which is the case that $\operatorname{End}_{\mathbf{C}} A = \mathbf{Z}$) or $\mathbf{B}[C_2]$. In the latter case, as explained in [Joh17, §4.3], there is a quadratic extension K/F, and a strongly irreducible weakly compatible system $\mathcal{S} = \{s_l\}$ of rank 2, weight 0 representations of G_K with determinant ε_l^{-1} such that $\mathcal{R}_A \cong \operatorname{Ind}_{G_K}^{G_F} \mathcal{S}$. Furthermore, writing $\operatorname{Gal}(K/F) = \{1, \sigma\}$, s_l^σ and s_l do not become isomorphic after restriction to any finite extension of K. (The case when K/F is totally real can be handled using potential automorphy theorems for GL_2, but our argument (at this point at least) does not need to distinguish between the various infinity types of K.)

Lemma 9.2.2. *If A/F is a challenging abelian surface, then for a density one set of primes l, $\overline{\rho}_{A,l}$ is vast and tidy.*

Proof. If $\operatorname{End}_{\mathbf{C}} A = \mathbf{Z}$, then, for all sufficiently large l, $\overline{\rho}_{A,l}(G_F) = \mathrm{GSp}_4(\mathbf{F}_l)$ by [Ser00], so the claim follows from Lemma 7.5.15.

If A is of type $\mathbf{B}[C_2]$, then writing $\rho_{A,l} \cong \operatorname{Ind}_{G_K}^{G_F} s_l$, we see from Lemma 7.5.22 and Lemma 9.1.10 (3) that we need only check that for a density one set of primes l, we have $\operatorname{Proj} \overline{s}_l^\sigma \not\cong \operatorname{Proj} \overline{s}_l$. (Note that the inverse of the mod l cyclotomic character is surjective for all l which are unramified in F.) To see this, note that since $\operatorname{Proj} s_l^\sigma \not\cong \operatorname{Proj} s_l$, we have $\operatorname{Sym}^2 s_l^\sigma \not\cong \operatorname{Sym}^2 s_l$ by [DK00, Appendix, Thm. B]. There is therefore some finite place v of F at which the compatible systems $\{\operatorname{Sym}^2 s_l^\sigma\}$, $\{\operatorname{Sym}^2 s_l\}$ are unramified, for which the eigenvalues of Frob_v differ for the two

compatible systems. Then the same applies for $\mathrm{Sym}^2 \overline{s}_l^\sigma$, $\mathrm{Sym}^2 \overline{s}_l$ for all sufficiently large l, so that in particular $\mathrm{Proj}\, \overline{s}_l^\sigma \not\equiv \mathrm{Proj}\, \overline{s}_l$, as required. □

Definition 9.2.3. Let A/F be an abelian surface over a totally real field. We say that a rational prime $p \geq 3$ is a *good prime for A* if:

- A admits a polarization of degree prime to p.
- p splits completely in F.
- The representation $\overline{\rho}_{A,p}$ is vast and tidy.
- For each place $v|p$, $\rho_{A,p}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary.

Remark 9.2.4. The point of Definition 9.2.3 is that the good primes p are the ones for which we can apply our modularity lifting theorem (Theorem 8.4.1) to $\rho_{A,p}$.

Lemma 9.2.5. *Let A/F be a challenging abelian surface. Then the set of rational primes which are good primes for A has relative density one in the set of primes which split completely in F.*

Proof. By Lemma 9.2.2, it suffices to show that $\overline{\rho}_{A,p}|_{G_{F_v}}$ is p-distinguished weight 2 ordinary for a density one set of finite places v of F (with residue characteristic p). To do this, we follow the approaches of [Saw16] and [CG20, Lem. A.7]. Consider the places v of F that are split over a prime p of $\mathbf{Q}$, for which A has good reduction; the set of such primes has density one. Fix a prime $l \neq p$. The characteristic polynomial of $\rho_{A,l}(\mathrm{Frob}_v)$ is of the form

$$x^4 - a_1 x^3 + a_2 x^2 - p a_1 x + p^2$$

where a_1, a_2 are integers.

Then A has good ordinary reduction at v if and only if $p \nmid a_2$. If this holds, then we see that $\overline{\rho}_{A,p}|_{G_{F_v}}$ will be p-distinguished weight 2 ordinary if and only if $a_1^2 - 4a_2$ is not divisible by p. By the Weil bounds, we have $|a_1| \leq 4\sqrt{p}$, $|a_2 - 2p| \leq 4p$, so if $a_1^2 - 4a_2$ is divisible by p, then it is equal to pc for c in some finite list of integers, independent of p.

Let G be the Zariski closure of $\rho_{A,l}(G_F)$ in GSp_4, and write V for the standard representation of GSp_4, and χ for the similitude character. Arguing exactly as in the proof of [Saw16, Thm. 1], it follows from the Cebotarev density theorem that it is enough to show that the virtual representation $(V^{\otimes 2} - 4 \wedge^2 V) \otimes \chi^{-1}$ does not have constant trace on any connected component of G.

By the proof of [Saw16, Thm. 3], we can replace G by the Sato–Tate group of A, which is either the connected group USp_4 (if A has type $\mathbf{A}$), or the normalizer of $\mathrm{SU}_2 \times \mathrm{SU}_2$ in USp_4 (if A has type $\mathbf{B}[C_2]$)) (which has two connected components). The result now follows easily from an explicit check. □

Lemma 9.2.6. *Let A/F be a challenging abelian surface. Then there are distinct rational primes p, q such that p and q are both good primes for A, and for all places $v|p$ of F, $\overline{\rho}_{A,q}(\mathrm{Frob}_v)$ has distinct eigenvalues.*

Proof. By Lemma 9.2.5, a density one subset of the set of rational primes which split completely in F are good primes. Let p be any good prime for A; then, for each place $v|p$ of F, $\rho_{A,p}|_{G_{F_v}}$ has p-distinguished weight 2 ordinary reduction, and in particular the eigenvalues of the crystalline Frobenius Frob_v on $T_p A$ are distinct. Consequently, for all but finitely many rational primes q of good reduction for A, $\overline{\rho}_{A,q}(\mathrm{Frob}_v)$ has distinct eigenvalues for all places $v|p$. □

If q is a good prime for an abelian surface A/F, then for each place $w|q$ of F we may write

$$\overline{\rho}_{A,q}|_{G_{F_w}} \cong \begin{pmatrix} \lambda_{\overline{\alpha}_w} & 0 & * & * \\ 0 & \lambda_{\overline{\beta}_w} & * & * \\ 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}_w}^{-1} & 0 \\ 0 & 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}_w}^{-1} \end{pmatrix}$$

Then we write

$$(\overline{\rho}_{A,q}|_{G_{F_w}})^{\mathrm{ss}} := \begin{pmatrix} \lambda_{\overline{\alpha}_w} & 0 & 0 & 0 \\ 0 & \lambda_{\overline{\beta}_w} & 0 & 0 \\ 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\beta}_w}^{-1} & 0 \\ 0 & 0 & 0 & \overline{\varepsilon}^{-1}\lambda_{\overline{\alpha}_w}^{-1} \end{pmatrix}$$

When reading the proofs of the following two results, it may be helpful to recall that our convention is that the representation $\overline{\rho}_{A,p}$ is the dual of $A[p]$; this accounts for the various duals occurring in the proofs.

Lemma 9.2.7. *Let A/F be an abelian surface over a totally real field, and let p, q be primes as in Lemma 9.2.6. Fix a totally real quadratic extension F_1/F in which p and q split completely, and which is linearly disjoint from the kernels of the actions of G_F on $A[p]$ and $A[q]$.*

Then there is a finite Galois extension of totally real fields F'/F, and a representation $\overline{r}_q : G_{F'F_1} \to \mathrm{GL}_2(\mathbf{F}_q)$, with the following properties:

(1) p and q both split completely in F'.
(2) F'/F is linearly disjoint from F_1/F and from the kernels of the actions of G_F on $A[p]$ and $A[q]$.
(3) $\det \overline{r}_q = \overline{\varepsilon}^{-1}$.
(4) $\overline{r}_q(G_{F'F_1}) = \mathrm{GL}_2(\mathbf{F}_q)$, and the projective image of $\overline{r}_q$ is not equal to its conjugate under $\mathrm{Gal}(F'F_1/F')$.
(5) Set $\overline{\rho}_q := \mathrm{Ind}_{G_{F'F_1}}^{G_{F'}} \overline{r}_q : G_{F'} \to \mathrm{GSp}_4(\mathbf{F}_q)$ with similitude factor $\overline{\varepsilon}^{-1}$. Then
   - *for any place $w|q$ of F and any place $w'|w$ of F', $\overline{\rho}_q|_{G_{F'_{w'}}} \cong (\overline{\rho}_{A,q}|_{G_{F_w}})^{\mathrm{ss}}$, and*
   - *for any place $v|p$ of F and any place $v'|v$ of F', $\overline{\rho}_q|_{G_{F'_{v'}}} \cong \overline{\rho}_{A,q}|_{G_{F_v}}$.*
(6) The representation $\overline{\rho}_q$ is vast and tidy.

Proof. Fix a finite place $\mathfrak{r}$ of F not dividing pq and splitting in F_1. We apply Proposition 9.1.12, taking $E = F$, $E' = F_1$, $G = \mathrm{GL}_2(\mathbf{F}_q)$, S to be the set of places dividing $pq\mathfrak{r}\infty$, and $F^{(\mathrm{avoid})}$ to be the extension cut out by the intersection of the kernels of $\overline{\rho}_{A,p}$ and $\overline{\rho}_{A,q}$. For each infinite place $v \in S'$ we choose c_v to have eigenvalues $\{1, -1\}$. For each place $w \in S$ dividing $pq\mathfrak{r}$ we write $w = w_1 w_2$ for its decomposition in F_1. If $w|p$, then the eigenvalues of $(A[q]^\vee|_{G_{F_w}})(\mathrm{Frob}_w)$ can be written as α_w, β_w, $p\beta_w^{-1}$, $p\alpha_w^{-1}$, while if $w|q$ we use the notation above. In either case, we choose $\overline{\phi}_{w_1}$ to correspond to the representation $\begin{pmatrix} \lambda_{\alpha_w} & 0 \\ 0 & \varepsilon^{-1}\lambda_{\alpha_w}^{-1} \end{pmatrix}$, and $\overline{\phi}_{w_2}$ to correspond to $\begin{pmatrix} \lambda_{\beta_w} & 0 \\ 0 & \varepsilon^{-1}\lambda_{\beta_w}^{-1} \end{pmatrix}$. Finally, if $w = \mathfrak{r}$, then we choose $\overline{\phi}_{w_1}$, $\overline{\phi}_{w_2}$ to have determinant $\overline{\varepsilon}^{-1}$, in such a way that $\overline{\phi}_{w_1}$ is unramified, while $\mathrm{Proj}\,\overline{\phi}_{w_2}$ is ramified.

We obtain an extension F'/F (the extension K/E from Proposition 9.1.12, with the ϕ_v there being our $\overline{\phi}_v$) and a representation $\overline{r}_q : G_{F_1F'} \to \mathrm{GL}_2(\mathbf{F}_q)$ which satisfies (1), and (2). It need not satisfy (3), but by construction $\overline{\varepsilon}\det\overline{r}_q$ is an even character which is trivial at all places dividing $pq\mathfrak{r}$. The obstruction to the existence of a square root of $\overline{\varepsilon}\det\overline{r}_q$ is therefore a class in the 2-torsion of $\mathrm{Br}_{F_1F'}$ which is trivial at all places dividing $pq\mathfrak{r}\infty$.

We can therefore replace F' by a quadratic totally real extension in which p, q, $\mathfrak{r}$ split completely, and assume that $\overline{\varepsilon}\det\overline{r}_q$ has a square root. By [AT09, Ch. X, Thm. 5] we can (by replacing this square root by a twist by a quadratic character) arrange that the square root is trivial at all places dividing $pq\mathfrak{r}$. Replacing $\overline{r}_q$ by its twist by this square root, we ensure (3), at which point (5) follows (note that for each place $v|p$ of F, $\overline{\rho}_{A,q}|_{G_{F_v}}$ is unramified with distinct eigenvalues of Frob_v, and is therefore semisimple). Considering the places lying over $\mathfrak{r}$, we see that (4) is satisfied. Finally, (6) then follows from Lemma 7.5.22. □

Theorem 9.2.8. *Let A/F be a challenging abelian surface over a totally real field. Then A is potentially modular. More precisely, there is a finite Galois extension of totally real fields F'/F and a prime p splitting completely in F' such that $\rho_{A,p}|_{G_{F'}}$ is modular and irreducible.*

Proof. Let p, q, F_1, F', $\overline{r}_q$ and $\overline{\rho}_q$ be as in Lemma 9.2.7. Let Y/F' denote the moduli space of triples $(B, \imath_p, \imath_q)$ consisting of abelian surfaces B and symplectic isomorphisms

$$\imath_p : B[p] \xrightarrow{\sim} A[p]|_{G_{F'}},$$
$$\imath_q : B[q] \xrightarrow{\sim} \overline{\rho}_q^\vee.$$

This is smooth and geometrically connected. (Over either $\mathbf{C}$ or $\overline{\mathbf{Q}}$, we may identify Y with the moduli space of principally polarized abelian surfaces with full level pq structure.)

We claim that for each place $v|pq\infty$ of F', the subspace $\Omega_v := Y^{\mathrm{ord}}(F'_v) \subset Y(F'_v)$ consisting of points corresponding to abelian surfaces with good ordinary reduction (when v is finite) is nonempty. If $v|\infty$, this follows from $\det r_q^\vee = \varepsilon$, while if $v|p$, then A itself gives a point of $Y(F'_v)$ (by point (5) of Lemma 9.2.7). Finally, if $v|q$, the canonical lift of A modulo v gives a point of $Y(F'_v)$. (Since A has good ordinary reduction at $v|p$ and $v|q$, the corresponding point on Y does indeed land in Ω_v.)

By [BLGGT14b, Prop. 3.1.1] (a theorem of Moret-Bailly), we may find a finite Galois totally real extension F''/F' in which p and q split completely, and which is linearly disjoint from the compositum of F_1F' and the kernels of the actions of $G_{F'}$ on $A[p]$, $A[q]$ and $\overline{\rho}_q$, with the property that $Y(F'') \cap \bigcap_{v|pq} \Omega_v \neq 0$. Let B/F'' be a corresponding abelian surface, which by construction will have good ordinary reduction for all $v|p$ and $v|q$.

By Proposition 9.1.11, after replacing F''/F' with a further totally real extension, we can maintain all of the above assumptions, and we can further suppose that there is a q-ordinary automorphic representation $\boldsymbol{\pi}$ of $\mathrm{GL}_2(\mathbf{A}_{F_1F''})$ of weight 0 and trivial central character, which is unramified at all places dividing pq and which satisfies $\overline{\rho}_{\boldsymbol{\pi},q} \cong \overline{r}_q|_{G_{F_1F''}}$.

It follows from [Rob01, Thm. 8.6] that there is an automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_{F''})$ of parallel weight 2 and trivial central character whose transfer to $\mathrm{GL}_4(\mathbf{A}_{F''})$ is the automorphic induction of $\boldsymbol{\pi} \otimes |\cdot|$, so that in particular $\rho_{\pi,q} \cong$

$\mathrm{Ind}_{G_{F_1F''}}^{G_{F''}} \rho_{\boldsymbol{\pi},q}$, so that $\overline{\rho}_{\pi,q} \cong \overline{\rho}_q|_{G_{F''}}$. In addition, π is ordinary, by construction. The representation $\rho_{\pi,p}$ is pure at all finite places because $\rho_{\boldsymbol{\pi},q}$ is (for the places away from q, this is proved in [Bla06], and for the places dividing q it is for example a very special case of the main theorem of [Car14]).

We can therefore apply Theorem 8.4.1 to $\rho_{B,q}$, and conclude that it is modular. Thus $\rho_{B,p}$ is modular, and applying Theorem 8.4.1 a second time, we deduce that $\rho_{A,p}|_{G_{F''}}$ is modular, as required. (The purity of $\rho_{B,q}, \rho_{B,p}$ and $\rho_{A,p}$ at all finite places is part of Proposition 2.8.1.) □

9.3. **Potential modularity and meromorphic continuation.** We now deduce the meromorphic continuation and functional equation of the L-functions associated to abelian surfaces over totally real fields from our potential modularity (and automorphy) results.

Theorem 9.3.1. *Let F be a totally real field, and let A/F be an abelian surface. Then $\mathcal{R}_A$ is potentially automorphic, and Conjecture 2.8.6 holds for A, for each $0 \leq i \leq 4$.*

Proof. Since $H^i(A, \mathbf{Q}_l) = \wedge^i H^1(A, \mathbf{Q}_l)$, it is enough to treat the cases $i = 1, 2$. Note that since for any $\mathcal{R}$ we have $\varepsilon(\mathcal{R})\varepsilon(\mathcal{R}^\vee) = N(\mathcal{R})$ (see [Tat79, (3.4.7)]), and we have $H^i(A, \mathbf{Q}_l)^\vee = H^i(A, \mathbf{Q}_l)(i)$, the claimed functional equation will follow from (2.8.5) in the case $\mathcal{R} = H^i(A, \mathbf{Q}_l)$.

To see that the meromorphic continuation and the functional equation (2.8.5) hold, note firstly that if A has type $\mathbf{D}$ or $\mathbf{F}$, then the compatible system $\mathcal{R}_A$ is potentially abelian, and the result follows from a standard argument with Brauer's theorem; more precisely, it is immediate from [Joh17, Prop. 11, Lem. 14]. In the general case, the same argument (see e.g. the proof of [Tay02, Cor. 2.2]) shows that it is enough to show that there is a Galois extension of totally real fields F'/F such that for each Galois extension F'/F'' with $\mathrm{Gal}(F'/F'')$ solvable, the compatible systems $\mathcal{R}_A|_{G_{F''}}$ and $\wedge^2\mathcal{R}_A|_{G_{F''}}$ are both automorphic. (Note that the meromorphic continuation and functional equations for the compatible systems follow from the functional equations (9.1.5) for the corresponding automorphic representations.)

Suppose now that A has type $\mathbf{B}[C_1]$, $\mathbf{C}$, or is of type $\mathbf{E}$ but not of type $\mathbf{E}[D_n]$. Then as we saw in the proof of 9.2.1, we can write $\mathcal{R}_A = \mathcal{R}_A^1 \oplus \mathcal{R}_A^2$ where $\mathcal{R}_A^1$, $\mathcal{R}_A^2$ are irreducible, odd, weight 0 weakly compatible systems of rank 2 l-adic representations of G_F. It follows from [BLGGT14b, Thm. 5.4.1] that there is a Galois extension of totally real fields F'/F such that $\mathcal{R}_A^1|_{G_{F'}}$, $\mathcal{R}_A^2|_{G_{F'}}$ are automorphic and irreducible. It follows from [BLGHT11, Lem. 1.3] that for each Galois extension F'/F'' with $\mathrm{Gal}(F'/F'')$ solvable, $\mathcal{R}_A^1|_{G_{F''}}$, $\mathcal{R}_A^2|_{G_{F''}}$ are automorphic. Thus $\mathcal{R}_A|_{G_{F''}}$ is automorphic, and since we have

$$\wedge^2\mathcal{R}_A|_{G_{F''}} = \det \mathcal{R}_A^1|_{G_{F''}} \oplus \det \mathcal{R}_A^2|_{G_{F''}} \oplus \left(\mathcal{R}_A^1|_{G_{F''}} \otimes \mathcal{R}_A^2|_{G_{F''}}\right),$$

it follows from [Ram00, Thm. M] that $\wedge^2\mathcal{R}_A|_{G_{F''}}$ is also automorphic, as required.

In the remaining cases, namely those of types $\mathbf{A}$, $\mathbf{B}[C_2]$, or $\mathbf{E}[D_n]$, it follows from Theorem 9.2.8, (the proof of) Proposition 9.2.1, Lemma 2.9.1 and Theorem 2.9.3, together with Lemma 8.3.2, that there is a Galois extension of totally real fields F'/F such that for some p, and all Galois extension F'/F'' with $\mathrm{Gal}(F'/F'')$ solvable, $\rho_{A,p}|_{G_{F''}}$ is irreducible and modular (and also automorphic). By the main result of [Hen09] (which is a refinement of the main result of [Kim03]), together with Theorem 2.9.3, we see that $\wedge^2\rho_{A,p}|_{G_{F''}}$ is automorphic, as required. □

Remark 9.3.2. Our use of the results of [Kim03] and [Hen09] in the proof of Theorem 9.3.1 is almost certainly overkill, and can be avoided by working with automorphic forms on GSp_4 rather than GL_4 as we now explain. Since the four 4-dimensional Galois representations $H^1(A, \mathbf{Q}_l)$ are generalized symplectic with respect to the Weil pairing, the exterior square $\wedge^2 H^1(A, \mathbf{Q}_l) = H^2(A, \mathbf{Q}_l)$ decomposes as the direct sum of a 5-dimensional Galois representation and the one dimensional summand $\mathbf{Q}_l(-1)$. (The corresponding Galois invariant classes in $H^2(A, \mathbf{Q}_l(1))$ are generated by the image of a hyperplane section under the cycle map.) Once one knows that the 4-dimensional representation $H^1(A, \mathbf{Q}_l)$ corresponds (potentially) to an automorphic representation π for GSp_4, then the L-function associated to the 5-dimensional summand of $H^2(A, \mathbf{Q}_l)$ is none other than the degree 5 standard L-function, whose analytic properties have been known for some time (see §6.3 of [GPSR87]). On the other hand, many of our arguments in this paper *do* crucially require passing between GSp_4 and GL_4 using Theorem 2.9.3 and Lemma 2.9.1. In particular, the proof of Theorem 9.3.1 uses base change in the form of Lemma 8.3.2, and therefore depends directly on Theorem 2.9.3; and of course our main modularity lifting theorems also depend on these results, in particular to prove that the modules that we patch are balanced.

If C/F is a curve over a number field, then we can define the completed L-functions $\Lambda_i(C, s)$ and the completed Hasse–Weil L-function $\Lambda(C, s)$ exactly as for abelian varieties. By definition we have $\Lambda_1(C, s) = \Lambda_1(\mathrm{Jac}(C), s)$, where $\mathrm{Jac}(C)$ is the Jacobian of C.

Corollary 9.3.3. *Let C/F be a genus two curve over a totally real field. Then the completed Hasse–Weil L-function $\Lambda(C, s)$ has a meromorphic continuation to the entire complex plane, and satisfies a functional equation of the form $\Lambda(C, s) = \varepsilon N^{-s} \Lambda(C, 3 - s)$ where $\varepsilon \in \mathbf{R}$ and $N \in \mathbf{Q}_{>0}$.*

Proof. This follows from Theorem 9.3.1 with $A = \mathrm{Jac}(C)$. □

Finally, we treat the case of genus one curves over quadratic extensions of totally real fields.

Theorem 9.3.4. *Let K/F be a quadratic extension of a totally real field F, and let E/K be either a genus one curve or an elliptic curve. Then E is potentially modular. More precisely, there is a Galois extension of totally real fields F'/F and a weight 0 cuspidal automorphic representation $\boldsymbol{\pi}$ of $\mathrm{GL}_2(\mathbf{A}_{KF'})$ with trivial central character such that for each prime l, we have $\rho_{E,l}|_{G_{KF'}} \cong \rho_{\boldsymbol{\pi},l}$, and in fact for each finite place v of KF' we have $\mathrm{WD}_v(\rho_{E,l}|_{G_{KF'_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}(\boldsymbol{\pi}_v |\det|_v^{-1/2})$.*

Furthermore, Conjecture 2.8.6 holds for E.

Proof. We may immediately replace E by its Jacobian and hence assume that E is an elliptic curve. If E is CM, then it is modular, while if E is isogenous to a twist of its Galois conjugate over F, then the result follows as in the proof of Proposition 9.2.1. We therefore assume that neither of these applies, and set $A = \mathrm{Res}_{K/F} E$. Then A is an abelian surface of type $\mathbf{B}[C_2]$, and $\mathcal{R}_A = \mathrm{Ind}_{G_K}^{G_F} \mathcal{R}_E$. By Theorem 9.2.8, there is a Galois extension of totally real fields F'/F, linearly disjoint from K/F, and an automorphic representation π of $\mathrm{GSp}_4(\mathbf{A}_{F'})$ such that $\rho_{A,p}|_{G_{F'}} \cong \rho_{\pi,p}$.

Let Π be the transfer of π to $\mathrm{GL}_4(\mathbf{A}_{F'})$. If κ is the quadratic character of $G_{F'}$ corresponding to $K' := KF'/F'$, it follows that $\Pi \otimes (\kappa \circ \mathrm{Art}_{F'} \circ \det) \cong \Pi$, so by [AC89,

Thm. 4.2, 5.1 of §3] there is a cuspidal automorphic representation $\boldsymbol{\pi}$ of $\mathrm{GL}_2(K')$ such that Π is the automorphic induction of $\boldsymbol{\pi} \otimes |\det|$. Write $\mathrm{Gal}(K'/F') = \{1, \tau\}$. We claim that $\boldsymbol{\pi}$ is of weight 0 and has trivial central character. Admitting this claim, it follows from Theorem 2.7.3 that we can write

$$\rho_{E,p}|_{G_{K'}} \oplus (\rho_{E,p}|_{G_{K'}})^\tau \cong \rho_{\boldsymbol{\pi},p} \oplus \rho_{\boldsymbol{\pi},p}^\tau$$

where all four 2-dimensional representations are irreducible. After possibly replacing $\boldsymbol{\pi}$ by $\boldsymbol{\pi}^\tau$, we conclude that $\rho_{E,p}|_{G_{KF'}} \cong \rho_{\boldsymbol{\pi},p}$, so that by Theorem 2.7.3, for each place $v \nmid p$ of K' we have $\mathrm{WD}_v(\rho_{E,l}|_{G_{KF'_v}})^{\mathrm{ss}} \cong \mathrm{rec}(\boldsymbol{\pi}_v |\det|_v^{-1/2})^{\mathrm{ss}}$. It follows that in fact

$$\mathrm{WD}_v(\rho_{E,l}|_{G_{K'_v}})^{F-\mathrm{ss}} \cong \mathrm{rec}(\boldsymbol{\pi}_v |\det|_v^{-1/2})$$

(because we know the corresponding statement for A). Repeating the argument for a second prime p, we see that this holds for all finite places v.

It remains to prove the claim. By Lemma 2.6.1, for each place $v|\infty$ of K', either $\boldsymbol{\pi}_v$ corresponds to $\phi_{0,1}$, or v is complex, and the L-parameter of $\boldsymbol{\pi}_v$ is scalar, given by $(z/\overline{z})^{\pm 1}$; in particular, in either case it is algebraic, and so the central character $\chi_{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ is algebraic. Moreover, if $\chi_{\boldsymbol{\pi}}$ is trivial, then the second case cannot occur, so that $\boldsymbol{\pi}$ automatically has weight 0. We therefore assume from now on that $\chi_{\boldsymbol{\pi}} \neq 1$, and derive a contradiction.

Since $\Pi^\vee \cong \Pi \otimes |\cdot|^{-2}$, we have

$$\boldsymbol{\pi} \boxplus \boldsymbol{\pi}^\tau \cong \boldsymbol{\pi}^\vee \boxplus (\boldsymbol{\pi}^\tau)^\vee,$$

so that either $\boldsymbol{\pi}^\vee \cong \boldsymbol{\pi}$, or $\boldsymbol{\pi}^\vee \cong \boldsymbol{\pi}^\tau$. In the former case, we would have $\boldsymbol{\pi} \cong \boldsymbol{\pi}^\vee \cong \boldsymbol{\pi} \boxtimes \chi_{\boldsymbol{\pi}}^{-1}$, from which it follows (if $\chi_{\boldsymbol{\pi}} \neq 1$) that $\chi_{\boldsymbol{\pi}}$ is the character of a quadratic extension L'/K' and $\boldsymbol{\pi}$ is induced from $\mathrm{GL}(1)/L'$. This implies that $\rho_{A,p}$ is potentially abelian, and thus that E is CM, a contradiction.

We can therefore assume that $\boldsymbol{\pi}^\vee \cong \boldsymbol{\pi}^\tau$, so that $\chi_{\boldsymbol{\pi}^\tau} = \chi_{\boldsymbol{\pi}}^{-1}$, and we shall derive a contradiction from these assumptions. Write χ for the p-adic character $G_{K'} \to \overline{\mathbf{Q}}_p^\times$ corresponding to the algebraic character $\chi_{\boldsymbol{\pi}}$. Let $v \nmid p$ be a place of K' for which $\rho_{A,p}|_{G_{K'}}$ is unramified, and let the eigenvalues of $\rho_{E,p}(\mathrm{Frob}_v)$ be $\{\alpha_v, q_v/\alpha_v\}$ and those of $(\rho_{E,p})^\tau(\mathrm{Frob}_v)$ be $\{\beta_v, q_v/\beta_v\}$. Now, $\alpha_v q_v^{-1/2}$ is either a Satake parameter of $\boldsymbol{\pi}$ or $\boldsymbol{\pi}^\tau$, so (using that $\chi_{\boldsymbol{\pi}^\tau} = \chi_{\boldsymbol{\pi}}^{-1}$) it follows that one of $\beta_v, q_v/\alpha_v, q_v/\beta_v$ is equal to either $q_v\chi(\mathrm{Frob}_v)/\alpha_v$ or $q_v\chi^{-1}(\mathrm{Frob}_v)/\alpha_v$.

Since χ is non-trivial, there is a set of places S of K' of positive density such that $\chi(\mathrm{Frob}_v) \neq 1$. Shrinking S if necessary, we deduce that there exists a set S of positive density so that one of the following equalities holds for all $v \in S$:

$$\begin{aligned} q_v\chi(\mathrm{Frob}_v)/\alpha_v &= \beta_v, \\ q_v\chi(\mathrm{Frob}_v)/\alpha_v &= q_v/\beta_v, \\ q_v\chi^{-1}(\mathrm{Frob}_v)/\alpha_v &= \beta_v, \\ q_v\chi^{-1}(\mathrm{Frob}_v)/\alpha_v &= q_v/\beta_v. \end{aligned}$$

By symmetry (replacing χ by χ^{-1} if necessary and β_v by q_v/β_v if necessary), we may assume that $\alpha_v\beta_v = q_v\chi(\mathrm{Frob}_v)$ for all $v \in S$.

Now, the representations $\rho_{E,p}|_{G_{K'}}$ and $(\rho_{E,p}|_{G_{K'}})^\tau$ have monodromy groups $\mathrm{GL}(2)$ by [Ser68, Thm. IV.2.2] and are not twist equivalent (by our running assumptions). It follows that the monodromy group of their tensor product is the identity component of $\mathrm{GO}(4)$. Since this is connected, we deduce by considering the formal

character of the corresponding Lie algebra $\mathfrak{sl}_2 \times \mathfrak{gl}_2$ that for any fixed character ξ, the generic element of the tensor product $\xi^{-1} \otimes \rho \otimes \rho^\tau$ does not have 1 has an eigenvalue.

However, for each place $v \in S$, we have (since $\chi(\mathrm{Frob}_v) = \alpha_v\beta_v/q_v$ for $v \in S$):

$$\varepsilon^{-1}(v)\chi^{-1}(v) \otimes \{\alpha_v, q_v/\alpha_v\} \otimes \{\beta_v, q_v/\beta_v\} = \{1, q_v^2/\alpha_v^2\beta_v^2, q_v/\beta_v^2, q_v\alpha_v^2\}.$$

Since S has positive density, this is a contradiction, as required. □

9.4. **K3 surfaces of large rank.** If A is an abelian surface over a totally real field F, then one may define the Kummer surface $\mathrm{Km}(A)$ to be the resolution of the quotient of A under the map $x \mapsto -x$. The variety $\mathrm{Km}(A)$ is a smooth projective algebraic K3 surface with (geometric) Picard number ≥ 17. (All Picard numbers in this section will be geometric Picard numbers.)

Proposition 9.4.1. *Let A be an abelian surface over a totally real field F, and let $X = \mathrm{Km}(A)$. Then Conjecture 1.1.1 holds for X.*

Proof. The cohomology groups $H^*(X, \mathbf{Q}_p)$ are trivial in odd degree. In even degree, they are generated by $H^*(A, \mathbf{Q}_p)$ plus the 16 dimensional space of Tate cycles in $H^2(X, \mathbf{Q}_p)$ spanned by the 16 exceptional divisors in the resolution $X \to A/(\pm 1)$. The latter classes are all defined over a finite extension of $\mathbf{Q}$, and hence the Galois representation (up to twist) they generate is an Artin representation. Hence the result follows from Theorem 9.3.1 applied to A, together with the meromorphic continuation of Artin L-functions. □

More generally, if a K3 surface X/F admits a Shioda–Inose structure [Mor84, §6] over F, then $H^*(X, \mathbf{Q}_p) \simeq H^*(\mathrm{Km}(A), \mathbf{Q}_p)$ for some abelian surface A/F, and Prop 9.4.1 implies Conjecture 1.1.1 for X. It might also happen that X/F admits a Shioda–Inose structure over some finite extension E. Recall Ribet's notion of a $\mathbf{Q}$-curve ([Rib04]) as an elliptic curve over $\overline{\mathbf{Q}}$ all of whose conjugates by $G_{\mathbf{Q}}$ are isogenous:

Definition 9.4.2. An F-abelian variety is an abelian variety A over a Galois extension E/F all of whose $\mathrm{Gal}(E/F)$-conjugates are isogenous to A over E.

Suppose that the conjugates A^σ over A are isogenous to (at most) quadratic twists of A (as necessarily happens if $\{\pm 1\}$ are the only automorphisms in $\mathrm{End}_{\mathbf{C}}(A)$). Then the Galois representations associated to $\wedge^2 H^1(A)$ and thus to $\mathrm{Km}(A)$ extend (even as compatible systems with $\mathbf{Q}$-coefficients) to G_F. Moreover, the (absolutely irreducible) projective Galois representations associated to $H^1(A)$ also extend to G_F, and thus, from the vanishing of $H^2(G_F, \mathbf{Q}/\mathbf{Z})$ due to Tate, also give rise to G_F representations (now with coefficients). If F is totally real and A is an F-abelian surface, one expects that the methods of this paper will have implications for the potential modularity of A. (Note that for primes p splitting completely in E, the mod p representations over F locally arise from abelian surfaces over $\mathbf{Q}_p$ — namely A itself.) We have not endeavored to undertake the task of proving results along these lines, however, since verifying that the Galois representations extend in the appropriate manner (especially when all the different possibilities for $\mathrm{End}_{\mathbf{C}}(A)$ are taken into account) would necessitate a somewhat involved analysis which we avoid due to issues of time and space.

9.4.3. *General K3 surfaces of Picard rank* ≥ 17. An algebraic K3 surface of Picard number 17 or 18 need not admit a Shioda–Inose structure even over $\mathbf{C}$ ([Mor84]). There need not even be a correspondence between X and an abelian surface A inducing a Hodge isometry of transcendental lattices $(T_X \otimes \mathbf{Q}) \simeq (T_A \otimes \mathbf{Q})$. The problem, as noted in [Mor84], is the following. Let U denote the hyperbolic plane — the lattice of rank two generated by two isotropic vectors which pair to 1. Then there are obstructions on the lattices T_X of signature $(2, 20-\rho(X)) = (2,3)$ or $(2,2)$ which arise from K3 surfaces to admit an injection of the form $(T_A \otimes \mathbf{Q}) \hookrightarrow (U \otimes \mathbf{Q})^3$. One might still hope to construct abelian varieties from K3 surfaces of large Picard rank by directly lifting the weight two polarized Hodge structure on T_X to a weight one Hodge structure of the smallest possible dimension. This amounts to considering the GSpin cover of the corresponding orthogonal group and relating that (in an ad hoc manner) to weight one Hodge structures via the identification of the associated Shimura variety as one of Hodge type. This differs slightly from the Kuga–Satake construction in which one has a functorial map from weight two Hodge structures to weight one Hodge structures via the Clifford algebra construction — the latter gives rise to abelian varieties in a uniform way, but introduces (in general) auxiliary dimensions, and, for a transcendental lattice T_X of rank 5, would produce an abelian variety of dimension $2^3 = 8$. In the case of interest to us, the corresponding GSpin Shimura variety will now (over $\mathbf{C}$) be precisely the moduli of abelian fourfolds with quaternionic multiplication (as considered in [KR99]), where the degenerate case $D = M_2(\mathbf{Q})$ corresponds to the usual moduli space of abelian surfaces. In particular, we arrive at the conclusion that a K3 surface with $\rho(X) = 17$ or 18 should either admit a correspondence with an abelian surface A inducing an isometry $(T_X \otimes \mathbf{Q}) \simeq (T_A \otimes \mathbf{Q})$, or there will exist an abelian fourfold A with quaternionic multiplication (a fake abelian surface, see the discussion after the statement of Lemma 10.3.2) and a correspondence inducing an injection $(T_X \otimes \mathbf{Q})^4 \hookrightarrow (T_A \otimes \mathbf{Q})$. This can also be predicted more arithmetically by using the Yoga of motives. For convenience, suppose that $\rho(X/F) = 17$. Let $\mathcal{R}$ be the compatible system associated to the transcendental motive (that is, the motive associated to the transcendental lattice), and assume that the Galois representations r_p are strongly irreducible for a density one set of primes p. One can try to lift $\mathcal{R}$ (up to quadratic twist) to a 4-dimensional compatible system $\mathcal{S}$ via the isogeny $\mathrm{GSp}_4 \to \mathrm{GO}_5$, and then realize $\mathcal{S}$ as the motive associated to an abelian surface. This happens, for example, when $X = \mathrm{Km}(A)$ for some A over F. In general, however, one encounters two obstructions. The first is that one should expect to have to extend coefficients of the motive by a compositum of quadratic fields. This is because a characteristic polynomial of an element in GO_5 with coefficients in $\mathbf{Q}$ lifts to a characteristic polynomial for GSp_4 whose coefficients lie either in $\mathbf{Q}$ or $\sqrt{D} \cdot \mathbf{Q}$ for some D. (This is an elementary computation with symmetric polynomials.) Let ${}^\sigma\mathcal{S}$ denote the compatible system obtained by applying an automorphism $\sigma \in G_{\mathbf{Q}}$ to the coefficients of $\mathcal{S}$. Since $\wedge^2\mathcal{S} = \mathcal{R} = \wedge^2({}^\sigma\mathcal{S})$, the irreducibility assumptions imply that $\mathcal{S}$ is a quadratic twist of ${}^\sigma\mathcal{S}$ for all $\sigma \in G_{\mathbf{Q}}$ acting on the coefficients. The restriction of $\mathcal{S}$ will thus have rational coefficients over some field E/F with $\mathrm{Gal}(E/F) = (\mathbf{Z}/2\mathbf{Z})^n$ where all the quadratic twists become trivial. If this restriction corresponds to an abelian surface A, this would predict (and even imply, see the remarks at end of [Mor84]) that there existed an algebraic cycle on $X \times \mathrm{Km}(A)$ which identified the corresponding transcendental

lattices over $\mathbf{Q}$. Moreover, the abelian surface A would be an F-abelian surface in the sense of Definition 9.4.2. On the other hand, even supposing $\mathcal{S}$ has $\mathbf{Q}$-coefficients over E, it need not be the case that $\mathcal{S}$ comes from an abelian surface, even though (for weight reasons) it must be an abelian motive. One also has to allow the possibility that it comes from a fake abelian surface, that is, a fourfold A with quaternionic multiplication (see the proof of Lemma 10.3.2). In summary, given a K3 surface X of Picard rank at least 17 over a number field F, one should be able to associate to X a canonical isogeny class of F-abelian surfaces or F-fake abelian surfaces. Under sufficiently big image hypotheses, it should be possible to rigorously justify the arguments of this paragraph using the methods and language of [Pat19, §4].

9.4.4. *Fake Kummer surfaces.* This raises the natural question as to whether, given an abelian fourfold with an inclusion $D \hookrightarrow \mathrm{End}^0(A)$ (a fake abelian surface), there are any natural geometrical constructions which produce a K3 surface (or, conversely, a construction in the other direction). For that matter, one might ask for an explicit geometric construction of either of these objects. Given six lines in general position in $\mathbf{P}^2$, the desingularization X of the double cover branched over those lines is, in general, a K3 surface of Picard rank 16. If the 6 lines are all tangent to a smooth conic, however, then the K3 surface generically has Picard rank 17, and moreover X is the Kummer surface associated to the Jacobian of the hyperelliptic curve obtained as the double cover of the conic branched at the six tangent points [Mor85]. This suggests looking for other degenerations of the six lines which could give rise to transcendental lattices with different integral structures.

The following construction, suggested to the authors by Madhav Nori, gives a $3 = 20 - 17$ dimensional rational family of such degenerations corresponding to $D = (-1, 3)_{\mathbf{Q}}$. Given a generic point in this family of Picard number 17, the corresponding $K3$ cannot be isogenous to a Kummer surface, and so indeed defines a genuine false Kummer surface. It is an interesting question to determine whether one can also see the corresponding abelian fourfold from this construction — possibly associated to a generalized Prym variety of some natural cover of curves under the map $\pi : X \to \mathbf{P}^2$. Consider five lines L_i for $i = 1, \dots 5$ in $\mathbf{P}^2$. These determine a conic C which passes through the intersections $L_1 \cap L_2$, $L_2 \cap L_3$, $L_3 \cap L_4$, $L_4 \cap L_5$, and $L_5 \cap L_1$, which we denote by P_i for $i = 1, \dots 5$. Let L_6 denote a sixth line which is tangent to C at P_6. Note that $C \cdot \sum_{i=1}^{6} L_i = 2\left(\sum_{i=1}^{6} P_i\right)$ is divisible by 2. Let Y denote the degree 2 cover of $\mathbf{P}^2$ and X its desingularization. The lifts of P_i in Y for $i = 1, \dots, 5$ are ordinary double points, and so the exceptional divisors E_i in X satisfy $E_i.E_i = -2$. Let $M = \sum_{i=1}^{5} E_i$. If $\pi : X \to \mathbf{P}^2$ denotes the projection, then $\pi^{-1}(C) = M + D$, where D is now an everywhere unramified double cover of C. But $C \simeq \mathbf{P}^1$, so D must decompose into two components $A + B$ meeting transversally at $\pi^{-1}(P_6)$. Note that $M.M = 5(-2) = -10$, that $A.B = 1$ (meeting transversally at $\pi^{-1}(P_6)$), and $\pi^{-1}(C).\pi^{-1}(C) = 2(C.C) = 8$. Moreover, $A.M = B.M = 5$, intersecting in E_i for $i = 1, \dots 5$. It follows that, if we let $E = A - B$, then $E.E$ is equal to

$$(A + B + M).(A + B + M) - 2(A + B).M - M.M - 4.A.B = 8 - 20 + 10 - 4 = -6.$$

The class E is transverse to all exceptional classes as well as the pre-image of the hyperplane class, so gives a new class in $\mathrm{NS}(X)$. Note that $U \otimes \mathbf{Q} \simeq (\langle 2k \rangle \oplus \langle -2k \rangle) \otimes \mathbf{Q}$ for any integer k. The transcendental lattice of the generic X is $(U^2 \oplus \langle -2 \rangle^2)_{\mathbf{Q}} \simeq$

$(\langle 6\rangle \oplus \langle -6\rangle \oplus U \oplus \langle -2\rangle^2)_{\mathbf{Q}}$, hence the corresponding transcendental lattice of this restricted family is rationally contained in $(U \oplus \langle -6\rangle \oplus \langle -2\rangle^2)_{\mathbf{Q}}$. Since this rational family has dimension $20-17=3$, the generic member will have Picard rank 17. As the form of the corresponding orthogonal group does not split, the lattice does not admit an injection into $(U\otimes\mathbf{Q})^3$, and so X is not isogenous to any Kummer surface. Indeed, from the rational structure of the resulting lattice, the corresponding fake abelian surface A will have endomorphisms by $D=(-1,3)_{\mathbf{Q}}$. (A related example was also considered in [LPS13] — in particular the divisor denoted in [LPS13] by X_6.)

9.4.5. *An Example.* Take the conic to be $y=x^2$, and the points P_i for $i=1,\dots 5$ to be (n,n^2) for $n=-2,\dots,2$. Now choose the point of tangency P_6 to be at $(3,9)$, so Y can be given by:

$$w^2=(-x+y)(x+y)(y-4z)(-3x+y+2z)(3x+y+2z)(-6x+y+9z).$$

The classes considered above are all defined over $\mathbf{Q}$, and so $\mathrm{Pic}(X/\mathbf{Q})\geq 17$. Let $\mathcal{R}=(\mathbf{Q},S,\{r_p\})$ denote the corresponding 5-dimensional compatible system of $\mathrm{GO}_5(\mathbf{Q}_p)$-representations. One checks that the set S of places of bad reduction is contained in the set of primes $\{p\leq 11,23,37,\infty\}$. Using both the determination of S and the fact that $\mathcal{R}(1)$ is self-dual, one computes that the determinant of $\mathcal{R}(1)$ is ψ, the quadratic character associated to $K=\mathbf{Q}(\sqrt{-2\cdot 3\cdot 7\cdot 23\cdot 37})$. The compatible system $\mathcal{R}(1)\otimes\psi$ is valued in $\mathrm{SO}_5(\mathbf{Q}_p)$. The corresponding symplectic compatible system $\mathcal{S}$ of rank 4 need not have coefficients in $\mathbf{Q}$, since it may come from a $\mathbf{Q}$-abelian variety in the sense of Definition 9.4.2. Indeed, it has coefficients in $\mathbf{Q}(\sqrt{3})$, and ${}^\sigma\mathcal{S}\simeq\mathcal{S}\otimes\psi$, where σ is the non-trivial element of $\mathrm{Gal}(\mathbf{Q}(\sqrt{3})/\mathbf{Q})$, and ψ is the character as above of conductor Δ_K. In particular, although A is a $\mathbf{Q}$-fourfold, the field of definition will be K.

Over $\mathbf{C}$, one expects that Nori's construction gives a rational parameterization of a component of the GSpin Shimura variety associated to the quaternion algebra $D=(-1,3)_{\mathbf{Q}}$ with some small (possibly trivial) level structure. Over $\mathbf{Q}$, the $\mathbf{Q}$-structure appears (by examining examples) to be associated to a twisted form associated to $\mathbf{Q}$-fourfolds A over a quadratic extension whose associated rank four motive over $\mathbf{Q}$ has coefficients in $\mathbf{Q}(\sqrt{3})$.

10. Applications to modularity

In this section, we apply our main modularity lifting theorem (Theorem 8.4.1) to prove modularity theorems for abelian surfaces. The methods generalize those of [Wil95, SBT97] for elliptic curves. In §10.3 and §10.4, we show that our results confirm the paramodular conjecture of [BK14] in many cases, but that there are counterexamples to the original formulation of the conjecture (arising from "fake abelian surfaces").

10.1. **First modularity results.** We begin this section with a proof of Theorem 1.1.7 of the introduction.

Proposition 10.1.1. *Let F be a totally real field in which $p>2$ splits completely. Let A/F be an abelian surface with good ordinary reduction at all places $v|p$, and suppose that at each $v|p$, the unit root crystalline eigenvalues are distinct modulo p. Assume that A admits a polarization of degree prime to p. Let*

$$\overline{\rho}_{A,p}:G_F\to \mathrm{GSp}_4(\mathbf{F}_p)$$

denote the dual of the mod p Galois representation associated to $A[p]$, and assume that $\overline{\rho}_{A,p}$ is vast and tidy. Assume that $\overline{\rho}_{A,p}$ is ordinarily modular, in the sense that there exists π of parallel weight 2 and central character $|\cdot|^2$ which is unramified and ordinary at all $v|p$, such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}_{A,p}$, and $\rho_{\pi,p}|_{G_{F_v}}$ is pure for all finite places v of F. Then A is modular. More precisely, there is an ordinary automorphic representation π' of $\mathrm{GSp}_4(\mathbf{A}_F)$ of parallel weight 2 and central character $|\cdot|^2$ which satisfies $\rho_{\pi',p} \cong \rho_{A,p}$.

Proof. As before, we write $\rho_{A,p} : G_F \to \mathrm{GSp}_4(\mathbf{Q}_p)$ for the Galois representation associated to the dual of the p-adic Tate module of A. The assumption that A admits a polarization of degree prime to p implies that the image of $\rho_{A,p}$ lands in $\mathrm{GSp}_4(\mathbf{Z}_p)$ and $\overline{\rho}_{A,p}$ lands in $\mathrm{GSp}_4(\mathbf{F}_p)$. By Proposition 2.8.1, the representation $\rho_{A,p}$ is pure for all places v of F. The assumption that A has good ordinary reduction for all $v|p$ and distinct unit root crystalline eigenvalues for all $v|p$ implies that the representations $\rho_{A,p}$ restricted to G_{F_v} are p-distinguished weight 2 ordinary. Prop 10.1.1 is then an immediate consequence of Theorem 8.4.1. □

Remark 10.1.2. If A does *not* have a polarization of order prime to p, then, by considering the kernel $A[\lambda]$ of any polarization $\lambda : A \to A^t$, we deduce that the representation $\overline{\rho}_{A,p} : G_{\mathbf{Q}} \to \mathrm{Aut}(A[p]) = \mathrm{GL}_4(\mathbf{F}_p)$ is reducible. Hence one could replace the assumption of the existence of a polarization on A of order prime to p in Prop. 10.1.1 by the assumption that the Galois representation associated to $A[p]$ is irreducible. On the other hand, we do not phrase our theorem in this way for the following reason: if A does not have a polarization of order prime to p, then it need not even be the case that the (necessarily reducible) representation $\overline{\rho}_{A,p} : G_{\mathbf{Q}} \to \mathrm{GL}_4(\mathbf{F}_p)$ associated to $A[p]$ lands in any conjugate of $\mathrm{GSp}_4(\mathbf{F}_p)$. Indeed, let $E/\mathbf{Q}$ be any elliptic curve such that $\overline{r}_{E,3} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_3)$ has surjective image, let $K/\mathbf{Q}$ be an auxiliary degree 3 cyclic extension, let $B = \mathrm{Res}_{K/\mathbf{Q}}(E)$, and let A denote the kernel of the map $B \to E$ induced from the trace map $\mathbf{Z}[\mathrm{Gal}(K/\mathbf{Q})] \to \mathbf{Z}$. Then A is an abelian surface, and $\overline{\rho}_{A,3} \simeq \overline{r}_{E,3} \otimes W$, where $W \in \mathrm{Ext}^1_{G_{\mathbf{Q}}}(\mathbf{F}_3, \mathbf{F}_3)$ is the unique non-trivial extension which splits over $\mathrm{Gal}(K/\mathbf{Q})$. The group theoretic image of $\overline{r}_{A,3}$ is isomorphic to $\mathrm{GL}_2(\mathbf{F}_3) \times \mathbf{Z}/3\mathbf{Z}$, but this is not isomorphic to any subgroup of $\mathrm{GSp}_4(\mathbf{F}_3)$. These examples are also related to the failure of the Shafarevich–Tate group Ш to have square order — William Stein [Ste04] found abelian surfaces A exactly of the form considered above with $3 \| Ш(A)[3^\infty]$.

We now give some examples where one can directly establish the modularity of certain residual representations.

Proposition 10.1.3. *Let F be a totally real field in which $p > 2$ splits completely. Let $\overline{\rho}_p : G_F \to \mathrm{GSp}_4(\mathbf{F}_p)$ be an absolutely irreducible representation with similitude factor ε^{-1} which is vast and tidy and p-distinguished weight 2 ordinary. Suppose furthermore that either:*

(1) $p = 3$, and $\overline{\rho}_3$ is induced from a 2-dimensional representation with inverse cyclotomic determinant over a totally real quadratic extension E/F in which 3 is unramified.
(2) $p = 5$, and $\overline{\rho}_5$ is induced from a 2-dimensional representation valued in $\mathrm{GL}_2(\mathbf{F}_5)$ with inverse cyclotomic determinant over a totally real quadratic extension E/F in which 5 is unramified.

(3) $\overline{\rho}_p$ is induced from a character of a quartic CM extension H/F in which p splits completely.

Then $\overline{\rho}_p$ is ordinarily modular, that is, there exists π of parallel weight 2 and central character $|\cdot|^2$ which is unramified and ordinary at all $v|p$, such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}$, and $\rho_{\pi,p}|_{G_{F_v}}$ is pure for all finite places v of F.

Proof. Suppose that we are in one of the first two settings, so that $p = 3$ or 5, and $\overline{\rho} = \mathrm{Ind}_{G_E}^{G_F} \overline{\varrho}$ for some representation $\overline{\varrho} : G_E \to \mathrm{GL}_2(\overline{\mathbf{F}}_p)$ with determinant ε^{-1}. The assumptions on $\overline{\rho}$ imply that $\overline{\varrho}|_{G_{E(\zeta_p)}}$ is absolutely irreducible, and the restriction of $\overline{\varrho}$ to the inertia group at any prime $w|p$ is an extension of ε^{-1} by 1. If $p = 5$, the condition on the determinant and the fact that E is unramified at p additionally ensures that the projective image of $\overline{\varrho}$ is not A_5. The representation $\overline{\varrho}$ locally has the structure of a representation associated to an ordinary Hilbert modular form of parallel weight two and trivial nebentypus. Suppose that $\overline{\varrho}$ is modular. It follows from [BLGG13, Thm. A] that $\overline{\varrho}$ does indeed arise from a Hilbert modular form of this kind, and we may take π to be the automorphic induction of this form from E to F. Since E/F is unramified, this will preserve the property of being ordinary. As in the proof of Theorem 9.2.8, purity follows from the main results of [Bla06, Car14]. Hence it suffices to establish the modularity of $\overline{\varrho}$.

If $\overline{\varrho}$ has solvable image, then, from a classification of the finite subgroups of $\mathrm{GL}_2(k)$ for a finite field k (see for example [SD73]), we deduce that the projective image of $\overline{\varrho}$ is either A_4, S_4, or dihedral, and is in particular a subgroup of $\mathrm{PGL}_2(\mathbf{C})$. By a theorem of Tate (see [Ser77, Theorem 4]), this implies that there exists a characteristic zero lift of $\overline{\varrho}$ which is totally odd with finite solvable image, and the result follows from an application of the theorems of Langlands and Tunnell ([Lan80, Tun81]) as in §5 of [Wil95]. It remains to consider the representations with vast non-solvable image. For $p = 3$, the only non-solvable induced representations which are vast come from representations (Lemma 7.5.21 (4)) $\overline{\varrho}_3 : G_F \to \mathrm{GL}_2(\mathbf{F}_9)$ with projective image A_5. The modularity of such a representation follows as in the solvable case, except now invoking the odd Artin conjecture for totally real fields ([PS16b, Thm. 0.3]) rather than Langlands–Tunnell. Alternatively, the arguments of [Ell05] over $\mathbf{Q}$ may be adapted to this setting.

Thus we are left with the case of non-solvable representations $\overline{\varrho} : G_E \to \mathrm{GL}_2(\mathbf{F}_5)$ with determinant ε^{-1}, which necessarily are surjective. The method of Khare–Wintenberger implies the existence of characteristic zero lifts of the required form (for example by [Sno09, Thm. 7.2.1] — the assumption that 5 is unramified in E guarantees that $[E(\zeta_5) : E] = 4$). To show that such a lift is modular, it suffices (by, for example, the main theorem of [Kis09]) to show that $\overline{\varrho}$ is modular. However, this follows from a standard argument going back to [SBT97, Tay03] by realizing $\overline{\varrho}$ as the 5-torsion of a modular elliptic curve over a solvable extension. In our situation, we may explicitly invoke [PS16b, Prop. 2.1.3].

Suppose finally that $\overline{\rho}_p = \mathrm{Ind}_{G_H}^{G_F} \chi$, where H/F is a quartic CM extension in which p splits completely. Let $v|p$ be a prime in F. The assumption that $\overline{\rho}_p$ is ordinary implies that for two of the primes $w|v$ of H the restriction of χ to inertia at w is ε^{-1}, and it is trivial at the other two primes above v. Let ψ denote an algebraic Grossencharacter of G_H with conductor prime to p and CM type corresponding to the mod-p weights of χ. If ψ_p is the p-adic avatar of ψ, then, by construction, the character $\psi_p/\chi \mod p$ is unramified at p, and hence, after

twisting ψ by the Teichmuller lift of this character, we may assume that $\psi_p \equiv \chi \mod p$. Let E/F denote the intermediate real quadratic field inside H. Then the automorphic induction of ψ to $\mathrm{GL}_2(\mathbf{A}_E)$ is a Hilbert modular form of parallel weight two which is ordinary at all $v|p$ and has trivial central character. Inducing once more to F, we obtain the required form π. □

For explicit examples of abelian surfaces $A/\mathbf{Q}$ with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$ whose mod-3 or mod-5 representations $\overline{\rho}_{A,p}$ satisfy Prop. 10.1.3 — and hence, by Prop. 10.1.1, are modular — see [CCG20]. In contrast to the examples found in [BPP+19] and [BK20] of large prime conductor, the examples found in [CCG20] have good reduction outside 2, 3, 5, and 7.

We also have the following application to modularity over number fields which need not be totally real (or even CM).

Theorem 10.1.4. *Let F be a totally real field in which 5 splits completely, and let K/F be a quadratic extension in which 5 is unramified. Let E/K be an elliptic curve which has good ordinary reduction or semistable ordinary reduction for all places $w|5$ of K. Finally, assume that the representation $\overline{\varrho}_{E,5} : G_K \to \mathrm{GL}_2(\mathbf{F}_5)$ has the following properties:*

(1) The projective image of $\overline{\varrho}_{E,5}$ is either $S_5 = \mathrm{PGL}_2(\mathbf{F}_5)$ or S_4.
(2) There exists a representation $\overline{r}_5 : G_F \to \mathrm{GL}_2(\mathbf{F}_5)$ with determinant ε^{-1} such that $\overline{r}_5|_{G_K} \cong \overline{\varrho}_{E,5}$.

Then E is modular. In particular, there exist infinitely many modular elliptic curves over K up to twist which are not CM and do not come from any subfield of K.

For example, one could take F to be $\mathbf{Q}(\sqrt{d})$ for any $d \equiv 1, 4 \mod 5$, and then take $K = \mathbf{Q}(\sqrt[4]{d})$, which is a field of mixed signature.

Proof. It suffices to prove that the twist of E by some quadratic character is modular. We now apply Lemma 7.5.26 to the representation $\overline{r}_5$ of G_F to obtain a representation

$$\overline{\rho} = \mathrm{Ind}_{G_K}^{G_F}(\overline{r}_5|_{G_K} \otimes \delta_{M/K}) = \mathrm{Ind}_{G_K}^{G_F}(\overline{\varrho}_{E,5} \otimes \delta_{M/K}),$$

where $\delta_{M/K}$ is the character of an auxiliary quadratic extension. By Lemma 7.5.27, this representation is vast and tidy and p-distinguished weight 2 ordinary.

As in the proof of Proposition 10.1.3, it follows from our hypotheses that $\overline{r}_5$ comes from an ordinary Hilbert modular form $\boldsymbol{\pi}$ for F. By taking the base change of this form to K/F, twisting by the quadratic character $\delta_{M/K}$, and then inducing back to F, we construct a π of parallel weight 2 and central character $|\cdot|^2$ which is unramified and ordinary at all $v|p$ such that $\overline{\rho}_{\pi,p} \cong \overline{\rho}$. Again, the purity of $\rho_{\pi,p}$ follows from the main results of [Bla06, Car14]. It follows from Theorem 8.4.1 that

$$\rho = \mathrm{Ind}_{G_K}^{G_F}(\varrho_{E,5} \otimes \delta_{M/K})$$

is modular, and hence (exactly as in the proof of Theorem 9.3.4) that $\varrho_{E,5} \otimes \delta_{M/K}$ and hence E is modular.

It is easy to produce examples of E satisfying the hypotheses of the theorem (starting with an elliptic curve over $\mathbf{Q}$, for example). Using the fact that the genus zero curve $X(\overline{\varrho}_{E,5})$ is isomorphic to $\mathbf{P}^1$ over K (there being at least one rational point coming from E), we deduce that there will be infinitely many such points. On the other hand, by choosing such points with appropriate local properties (for

example, ramified at one prime w above v but not at the other) we may find infinitely many examples which do not arise via base change. Since the mod 5-representations associated to these curves are not projectively dihedral or cyclic, they also cannot have CM. □

10.2. **Abelian varieties with fixed 3-torsion.** We have produced a number of residual representations mod p for small p which are automorphic. It is natural to ask whether any such representation (satisfying necessary local conditions) arises from infinitely many abelian surfaces over F. The corresponding question for 2-dimensional representations has a positive answer precisely when $p = 2$, 3, or 5, where the corresponding moduli space is a smooth curve of genus zero. We show that for abelian surfaces there is a positive answer for $p = 2$ and $p = 3$. When $p \geq 5$, the moduli space in question is of general type [HS02], and so one would not expect (in general) that they admit infinitely many rational points not lying on a special Shimura subvariety, although we do not attempt to address this question.

When $p = 2$, the problem is pretty much obvious. The fact that the corresponding moduli space for the trivial representation $\overline{\rho}$ is rational goes back to Igusa (see [HS02, Theorem IV.1.4] and [Igu64]). The fact that the corresponding moduli space for non-trivial $\overline{\rho}$ is unirational is also surely well-known (we shall now give a sketch of this result although we shall never use this fact). Fix a representation $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_2)$. Since $\mathrm{GSp}_4(\mathbf{F}_2) \simeq S_6$, one may write any G-extension L of F for $G \subseteq S_6$ as the splitting field of a degree 6 separable polynomial $f(x)$ over F. If one then takes A to be the Jacobian of the curve $y^2 = g(x)$ for any $g(x)$ with $\mathbf{Q}[x]/g(x) \simeq \mathbf{Q}[x]/f(x)$, then $\overline{\rho}$ is the representation associated to the 2-torsion of A. An elementary computation shows that this gives a $3 = 6 - \dim \mathrm{PGL}_2$ dimensional family of abelian surfaces up to isomorphism with fixed $\overline{\rho}$ for any such $\overline{\rho}$. Explicitly, one may let e_i for $i = 1$ to 6 be any basis over $\mathbf{Q}$ of the étale $\mathbf{Q}$-algebra $\mathbf{Q}[x]/f(x)$, and then let $g(x)$ be the minimal polynomial of $\sum t_i e_i$. (The Jacobian A depends only on $g(x)$ up to the action of PGL_2 on $\mathbf{P}^1$.)

This leaves the case $p = 3$. The answer in this case can be extracted from the very extensive literature on the subject, essentially following the main idea of [SBT97]. Fix $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_3)$ with inverse cyclotomic similitude character, and let V denote the underling symplectic space over $\mathbf{F}_3$. Let $B(\overline{\rho})/F$ denote the moduli space of pairs $(A, \imath_3)$ consisting of abelian surfaces A and symplectic isomorphisms

$$\imath_3 : A[3] \xrightarrow{\sim} V^\vee.$$

(The dual is here because our Galois representations have been normalized cohomologically, so it is the dual representations which actually occur inside the p-adic Tate modules.) The variety $B(\overline{\rho})$ is smooth and geometrically connected. Over $\mathbf{C}$, we may identify B with the moduli space of principally polarized abelian surfaces with full level 3 structure. This space is well-known to be a (geometrically) rational threefold, and is isomorphic to an open subvariety of the Burkhardt quartic [Bur91, Cob06, Bak46, Hun96, BN18], specifically, the complement in the Burkhardt quartic of the Hessian hypersurface.

The Burkhardt quartic is exceptional for a number of different reasons, not least of which is that it admits an action of the group $\mathrm{PSp}_4(\mathbf{F}_3)$ (tautologically from the description above). If $V = (\mu_3)^2 \oplus (\mathbf{Z}/3\mathbf{Z})^2$, we write B for $B(\overline{\rho})$. One knows ([BN18]) that B is rational over $\mathbf{Q}$. Suppose we knew that $\overline{\rho}$ actually came from an abelian surface A, so that $B(\overline{\rho})$ admitted a smooth rational point over F. One

might ask whether this is enough to force the twist $B(\overline{\rho})$ to be rational over F; this question is resolved in the negative in [CC20]. The difficulty in a naïve attempt to replicate the argument of Taylor and Shepherd-Barron ([SBT97]) in this case is that the birational map $B \to \mathbf{P}^3$ is *not* equivariant with respect to $\mathrm{PSp}_4(\mathbf{F}_3)$ and any embedding $\mathrm{PSp}_4(\mathbf{F}_3) \to \mathrm{PGL}_4(\overline{\mathbf{Q}}) = \mathrm{Aut}(\mathbf{P}^3)$. This means that a $\mathrm{PSp}_4(\mathbf{F}_3)$-twist of B does not naturally inherit the structure of a Severi–Brauer variety.

It turns out, however, that we are lucky. There exists a cover $P(\overline{\rho}) \to B(\overline{\rho})$ of degree 6 corresponding to an additional choice of level 2 structure of A, namely an odd theta characteristic, (or, for $A = \mathrm{Jac}(C)$, a Weierstrass point on the corresponding genus two curve C). The cover $P(\overline{\rho})$ now *does* have the property that it is not only rational, but $\mathrm{PSp}_4(\mathbf{F}_3)$-equivariantly rational, which allows us to deduce the rationality of $P(\overline{\rho})$ in favourable circumstances, and hence the unirationality of $B(\overline{\rho})$. In particular, this allows us to construct infinitely many rational points on $B(\overline{\rho})$ which correspond to abelian surfaces A with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$, as in the following theorem.

Theorem 10.2.1. *Fix $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_3)$ with similitude character ε^{-1}. Then $B(\overline{\rho})$ is unirational over F, and there exist infinitely many principally polarized abelian surfaces A/F up to twist with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$ and such that $\imath_3 : A[3] \simeq V(\overline{\rho})^\vee$. Moreover, we may additionally assume that these A are Jacobians of curves which have a rational Weierstrass point, and may thus be written in the form $y^2 = f(x)$ where $f(x)$ is a quintic polynomial. Suppose, in addition, that for all $v|3$, the representation $\overline{\rho}|_{G_{F_v}}$ arises as the 3-torsion of an abelian surface over F_v with good ordinary reduction. Then we may additionally assume that these A also all have good ordinary reduction for all $v|3$.*

Note that, as with of [SBT97, Thm. 1.2], we do not need to impose any further local hypotheses at any primes after we impose the global condition on the similitude character. (In particular, if K is a local field of characteristic zero, then by using a globalization argument as in [Cal12, Thm. 3.1], this implies that the only requirement on a mod 3-representation $\overline{\rho} : G_K \to \mathrm{GSp}_4(\mathbf{F}_3)$ to arise from a principally polarized abelian surface over K is that the similitude character is ε^{-1}.)

The remainder of this section is devoted to the proof of Theorem 10.2.1. We start by defining the non-Galois degree 6 cover P of B and recalling its basic properties.

Definition 10.2.2. Let $B(2) := \mathcal{A}_2(6) \to \mathcal{A}_2(3) = B$ denote the cover of B corresponding to a choice of full level-2 structure. It is a Galois cover with Galois group $S_6 \simeq \mathrm{PSp}_4(\mathbf{F}_2)$, where we fix this identification up to conjugacy by identifying S_6 generically with the Galois group of the Weierstrass points on C with $\mathrm{Jac}(C) = A$. (This identification can be made explicit using the map τ below.) Then P denotes the intermediate cover over B corresponding to the conjugacy class of subgroups $S_5 \subset S_6 = \mathrm{PSp}_4(\mathbf{F}_2)$ which fix a point.

A more natural definition of P is given in terms of theta characteristics. Namely, P may be identified with the moduli space $\mathcal{A}_2(3)^-$ of principally polarized abelian surfaces with a symmetric odd theta structure of level 3 (see [DL08, §2.3]). Recall that there is a natural Torelli map $\tau : \mathcal{M}_2(3) \to \mathcal{A}_2(3)$ which is a bijection away from the Humbert surface consisting of principally polarized abelian surfaces which split as the product of two elliptic curves. A rational point on P in the image of τ corresponds to $A = \mathrm{Jac}(C)$, together with a symplectic isomorphism from $A[3]$ to $(\mu_3)^2 \oplus (\mathbf{Z}/3\mathbf{Z})^2$ and the data of a rational Weierstrass point on C, which (after

moving this point to infinity) means that C can be written in the form $y^2 = f(x)$ for a quintic polynomial $f(x)$.

The level 3-structure on B pulled back to P gives an action over $\mathbf{Q}(\sqrt{-3})$ of the group $\mathrm{PSp}_4(\mathbf{F}_3)$. If one gives $\mathrm{PSp}_4(\mathbf{F}_3)$ the structure of an étale group scheme over $\mathbf{Q}$ by viewing it as the group G of symplectic automorphisms of $\mu_3^2 \times (\mathbf{Z}/3\mathbf{Z})^2$ modulo $(-1,-1)$, then this action descends to $\mathbf{Q}$. Equivalently, the twisting of any automorphism in $\mathrm{PSp}_4(\mathbf{F}_3)$ by the action of $\sigma \in G_{\mathbf{Q}}$ is accounted for by the Galois action on $\mu_3^2 \times (\mathbf{Z}/3\mathbf{Z})^2$ (as in the formula for $\sigma\theta$ below).

Proposition 10.2.3. *The variety P is rational over $\mathbf{Q}$. Moreover, there exists a birational map $P \to \mathbf{P}^3$ over $\mathbf{Q}$ which is equivariant with respect to the action of G for some action of G on $\mathbf{P}^3$.*

Proof. The G-equivariant map $P \to \mathbf{P}^3_{\mathbf{Q}}$ is the odd theta map denoted Th^- in §2.4 of [DL08]. The fact that Th^- is a birational isomorphism is [Bol07, Theorem 0.0.1]. □

We now turn to the proof of Theorem 10.2.1. From Proposition 10.2.3, it follows that the rationality of $P(\overline{\rho})$ over F is equivalent to the rationality of $\mathbf{P}^3(\overline{\rho})$ over F, where $\mathbf{P}^3(\overline{\rho})$ is the twist of $\mathbf{P}^3$ arising from the projective representation associated to $\overline{\rho}$. The action of $\mathrm{Sp}_4(\mathbf{F}_3)$ on $\mu_3^2 \times (\mathbf{Z}/3\mathbf{Z})^2$ over $\mathbf{Q}(\zeta_3)$ induces a homomorphism θ from $\mathrm{Sp}_4(\mathbf{F}_3)$ to $\mathrm{Aut}(P)$ and hence to $\mathrm{Aut}(\mathbf{P}^3)$. This map satisfies

$$\sigma\theta(\alpha) = \theta\left(\begin{pmatrix} \varepsilon(\sigma) & 0 & 0 & 0 \\ 0 & \varepsilon(\sigma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \alpha \begin{pmatrix} \varepsilon^{-1}(\sigma) & 0 & 0 & 0 \\ 0 & \varepsilon^{-1}(\sigma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right).$$

Since $\overline{\rho}$ has similitude factor $\overline{\varepsilon}^{-1}$, we can associate to $\overline{\rho}$ a cocycle

$$\sigma \mapsto \theta\left(\overline{\rho}^{\vee}(\sigma) \begin{pmatrix} \varepsilon^{-1}(\sigma) & 0 & 0 & 0 \\ 0 & \varepsilon^{-1}(\sigma) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right),$$

in $H^1(F, \mathrm{PGL}_4(\overline{\mathbf{Q}}))$, and $\mathbf{P}^3(\overline{\rho})$ is the twist of $\mathbf{P}^3$ by this cocycle.

Lemma 10.2.4. *If $\overline{\rho} : G_F \to \mathrm{GSp}_4(\mathbf{F}_3)$ has similitude factor ε^{-1}, then $P(\overline{\rho})$ is rational.*

Proof. The proof is very similar to the proof of [SBT97, Lem. 1.1]. We need to show that the cocycle in $H^1(F, \mathrm{PGL}_4(\overline{\mathbf{Q}}))$ corresponding to $\overline{\rho}$ vanishes, so it is enough to show that it comes from $H^1(F, \mathrm{GL}_4(\overline{\mathbf{Q}}))$. It is therefore enough (following the argument in [SBT97]) to show that we can lift the induced homomorphism $\theta : \mathrm{PSp}_4(\mathbf{F}_3) \to \mathrm{PGL}_4(\overline{\mathbf{Q}})$ to a *unique* homomorphism $\widetilde{\theta} : \mathrm{Sp}_4(\mathbf{F}_3) \to \mathrm{GL}_4(\overline{\mathbf{Q}})$. Since $\mathrm{PSp}_4(\mathbf{F}_3)$ is perfect (indeed simple) it has a unique Schur cover (Darstellungsgruppe). Since the Schur multiplier of $\mathrm{PSp}_4(\mathbf{F}_3)$ has order 2 ([CCN+85]), the Darstellungsgruppe of $\mathrm{PSp}_4(\mathbf{F}_3)$ may be identified with $\mathrm{Sp}_4(\mathbf{F}_3)$, and in particular the projective representation θ lifts to a genuine homomorphism $\widetilde{\theta} : \mathrm{Sp}_4(\mathbf{F}_3) \to \mathrm{GL}_4(\overline{\mathbf{Q}})$. It remains to show this lift is unique.

We claim this follows from the fact that $\mathrm{Sp}_4(\mathbf{F}_3)$ is perfect. Indeed, because the group is perfect, every element of $\mathrm{Sp}_4(\mathbf{F}_3)$ can be written as a product of commutators $[g,h]$. Hence it suffices to show that $\widetilde{\theta}([g,h])$ is uniquely defined.

But $\widetilde{\theta}([g,h]) = [\widetilde{\theta}(g), \widetilde{\theta}(h)]$, and the commutator of any two elements of $\mathrm{GL}_n(\overline{\mathbf{Q}})$ depends only on their images in $\mathrm{PGL}_n(\overline{\mathbf{Q}})$, as required. We note (although we do not use this fact) that $\mathrm{PSp}_4(\mathbf{F}_3)$ has no faithful 4-dimensional representations (again by [CCN+85]) and so the representation $\widetilde{\theta}$ of $\mathrm{Sp}_4(\mathbf{F}_3)$ is faithful. □

Remark 10.2.5. Note that the corresponding facts (that $\mathrm{PSL}_2(\mathbf{F}_5)$ is simple with Darstellungsgruppe $\mathrm{SL}_2(\mathbf{F}_5)$) lead to a proof of [SBT97, Lem. 1.1]. This differs slightly from the original proof in [SBT97] as follows: Instead of using the fact that $\mathrm{PSL}_2(\mathbf{F}_5)$ has Schur multiplier $\mathbf{Z}/2\mathbf{Z}$ and deducing that any irreducible projective representation θ lifts to a representation $\widetilde{\theta} : \mathrm{SL}_2(\mathbf{F}_5) \to \mathrm{GL}_2(\overline{\mathbf{Q}})$, the authors use the fact that the kernel of $\mathrm{SL}_2(\overline{\mathbf{Q}}) \to \mathrm{PSL}_2(\overline{\mathbf{Q}})$ has order 2, and so θ *automatically* lifts to a representation of a degree 2 central extension of $\mathrm{PSL}_2(\mathbf{F}_5)$ to $\mathrm{SL}_2(\overline{\mathbf{Q}})$, and then argue that the image (and hence source) is $\mathrm{SL}_2(\mathbf{F}_5)$ (because the split central extension would give a faithful 2-dimensional representation of $\mathrm{PSL}_2(\mathbf{F}_5)$).

We have now proved under the given hypothesis on $\overline{\rho}$ that $P(\overline{\rho})$ is rational, and hence $B(\overline{\rho})$ is unirational. Moreover, twisting P by $\overline{\rho}$ leaves the level structure at the odd theta characteristic unchanged, so that all the corresponding abelian varieties in the image of the Torelli map are Jacobians of curves C of the form $y^2 = f(x)$ where $f(x)$ is a quintic (after moving the rational Weierstrass point to infinity).

Proof of Theorem 10.2.1. We need to show that infinitely many of the corresponding points of $P(\overline{\rho})$ do not admit any extra endomorphisms over $\mathbf{C}$. We show that we may find infinitely many A such that the Galois representation associated to $A[5]$ has image containing $\mathrm{Sp}_4(\mathbf{F}_5)$. If A/F *did* admit extra endomorphisms over $\overline{F}$, then, from the classification of the possible Galois types of endomorphism structures on A recalled at the beginning of §9.2, the Galois representation associated to the 5-adic Tate module of A would become reducible after making an extension of degree at most 2. But the action of $\mathrm{Sp}_4(\mathbf{F}_5)$ on $\mathbf{F}_5^4$ remains absolutely irreducible after restriction to any index two subgroup, which forces $\mathrm{End}_{\overline{\mathbf{Q}}}(A) = \mathbf{Z}$. (In this argument, 5 could have been replaced by any prime p independent of the level structure.)

We will now arrange this condition by an application of Hilbert irreducibility as in the proof of [SBT97, Theorem 1.2]. Let $R(\overline{\rho}) \to P(\overline{\rho})$ be the fibre product of $P(\overline{\rho})$ with $\mathcal{A}_2(5) \to \mathcal{A}_2$. This is a Galois cover with Galois group $\mathrm{PSp}_4(\mathbf{F}_5)$. Recall that $P(\overline{\rho})$ is rational. By Hilbert irreducibility ([Ser89, §9.2, 9.6]), we may find infinitely many points $x \in P(\overline{\rho})(F)$ so that the Galois group of the splitting field of any preimage $y \in R(\overline{\rho})$ of x contains $\mathrm{PSp}_4(\mathbf{F}_5)$, and moreover, we may restrict x to any non-trivial open subset of $P(\overline{\rho})(F_v)$ for all v in some finite set of primes S. If A denotes the corresponding abelian surface, it follows that the projective Galois representation associated to $\overline{\rho}_{A,5}$ contains $\mathrm{PSp}_4(\mathbf{F}_5)$, and thus the image of $\overline{\rho}_{A,5}$ itself contains $\mathrm{Sp}_4(\mathbf{F}_5)$. If we now use the assumption that $P(\overline{\rho})(F_v)$ has points corresponding to abelian surfaces with good ordinary reduction at all $v|3$, then (since the ordinary condition is open) we can choose our $x \in P(\overline{\rho})(F)$ so that A/F has good ordinary reduction for all $v|3$. □

We obtain the following corollary:

Theorem 10.2.6. *Let F be a totally real field in which 3 splits completely. Then there exist infinitely many abelian surfaces A/F up to twist with $\mathrm{End}_{\mathbf{C}}(A) = \mathbf{Z}$ which are modular, and which do not come from any proper subfield of F (in the*

sense that, for each p, there is no twist of the corresponding Galois representation $\rho_{A,p}$ which extends to the absolute Galois group of a proper subfield of F).

Proof. Let H/F be a quadratic extension in which every prime $v|3$ is inert, and let σ be the non-trivial element of $\mathrm{Gal}(H/F)$. Let E/H be an elliptic curve with good ordinary reduction for all $v|3$, such that $\overline{\rho}_{E,3} : G_H \to \mathrm{GL}_2(\mathbf{F}_3)$ is surjective, and such that the projective images of $\overline{\rho}_{E,3}$ and $\overline{\rho}^{\sigma}_{E,3}$ are totally disjoint. Since $X(1)$ has genus zero, this can be achieved by choosing a global point which lies over a suitable choice of smooth point in $X(1)(k_w)$ (with $k_w = \mathcal{O}_H/w$) for suitably chosen primes w which split over F, for example ensuring that $\overline{\rho}_{E,3}(\mathrm{Frob}_w)$ and $\overline{\rho}_{E,3}(\mathrm{Frob}_{w^\sigma})$ give distinct elements of $\mathrm{PGL}_2(\mathbf{F}_3)$. Similarly, by making choices above primes of $\mathbf{Q}$ which split completely in H, we may ensure that $\overline{\rho} = \mathrm{Ind}_{G_H}^{G_F} \overline{\rho}_{E,3}$ does not descend after twisting to any proper subfield of F.

By Lemma 7.5.22, $\overline{\rho}$ is vast and tidy. The fact that each prime $v|3$ in H/F is inert implies that $\overline{\rho}$ is 3-distinguished and finite flat, and hence 3-distinguished weight 2 ordinary. It follows from Proposition 10.1.3 that $\overline{\rho}$ is ordinarily modular. The representation $\overline{\rho}$ at each $v|3$ arises locally from an abelian variety over F_v with good ordinary reduction, because it does so globally — namely, the restriction of scalars of E from H_v to F_v. It follows from Theorem 10.2.1 that there are infinitely many abelian surfaces A with up to twist with good ordinary reduction at each place $v|3$ and satisfying $A[3] \cong \overline{\rho}$. The choice of $\overline{\rho}$ ensures that any such A does not descend (even after twist) to any subfield of F. Finally, every such A is modular by Proposition 10.1.1. □

10.3. **The Paramodular Conjecture.** We end this section with a discussion of the relationship between our results and the "paramodular conjecture" of [BK14] (cf. also the remarks in [Yos80, §8, p.243]). Recall that this conjecture states that there should be a bijection (determined by the compatibility of Frobenius eigenvalues and Hecke eigenvalues at unramified places) between isogeny classes of abelian surfaces $A/\mathbf{Q}$ of conductor N with $\mathrm{End}_{\mathbf{Q}} A = \mathbf{Z}$, and holomorphic cuspidal Siegel newforms of weight 2 and paramodular level N which are "non-lifts" and have rational Hecke eigenvalues, considered up to scalar multiplication. Here "non-lifts" means that they are orthogonal to the space of Gritsenko lifts. We explain in this section why the paramodular conjecture as originally formulated in [BK14] is not true. The issue is that Siegel newforms of weight 2 and paramodular level N with rational eigenvalues will *not* always correspond to abelian surfaces. In light of the observations of this paper, Brumer and Kramer have modified their conjecture in [BK19] along the lines suggested by the analysis presented here — we reproduce their updated conjecture in this paper as Conjecture 10.4.3 below. In order to distinguish between the two versions of this conjecture, we refer to the original formulation (given above) as the *original* paramodular conjecture, and the modified version (Conjecture 10.4.3) as the paramodular conjecture. Both of these conjectures posit an injective map from isogeny classes of abelian surfaces $A/\mathbf{Q}$ of conductor N with $\mathrm{End}_{\mathbf{Q}} A = \mathbf{Z}$ to Siegel newforms of weight 2 and paramodular level N, which are "non-lifts" and have rational Hecke eigenvalues, and hence, when talking about the implication in this direction, we do not distinguish between different versions of the conjecture.

We firstly show that all of our examples of modular abelian varieties verify the paramodular conjecture, before giving a more general explanation of the relationship between the paramodular conjecture and the Langlands program, and then explaining some counterexamples to the original paramodular conjecture.

Lemma 10.3.1. *Any abelian variety $A/\mathbf{Q}$ satisfying the hypotheses of Proposition 10.1.1 satisfies the paramodular conjecture; that is, there is a corresponding holomorphic cuspidal Siegel newform of weight $(2,2)$ and paramodular level equal to the conductor of A, which is a non-lift, has rational Hecke eigenvalues, and is unique up to scalars.*

Proof. By Proposition 10.1.1 (or more precisely by Theorem 8.4.1, as applied in the proof of Proposition 10.1.1) there is an L-packet of cuspidal automorphic representations π of weight 2 and general type corresponding to A, whose L-parameters coincide with those determined by A. The claim that there is a unique corresponding newform of level equal to the conductor of A is now a consequence of the theory of newforms due to Roberts and Schmidt [RS07b] (which assumes that we are working with representations of trivial central character, but this is harmless, as we can reduce to this case by twisting π by $|\cdot|$). This newform is certainly a non-lift, as π is of general type (see the discussion following this lemma for a more precise description of the non-lifts), and it has rational Hecke eigenvalues by local-global compatibility.

More precisely, by [Sch18, Thm. 1.1], for each prime v of $\mathbf{Q}$, there is a unique paramodular representation in the L-packet at v, namely the unique generic representation. Since representations of general type are stable, this gives rise to a unique π of weight 2 which has a paramodular vector at each finite place. Furthermore, for each v the space of paramodular vectors at minimal paramodular level is one-dimensional by [RS07b, Thm. 7.5.1], and this minimal paramodular level coincides with the conductor of the corresponding L-parameter (and thus with that of A) by [RS07b, Thm. 7.5.4(iii)] and the main theorem of [GT11a]. □

We now discuss the paramodular conjecture more broadly. Firstly, we discuss the automorphic side of the conjecture. As explained in [Sch18], the space of Siegel modular forms of weight 2 and fixed level can be written as an orthogonal sum of spaces spanned by eigenforms in automorphic representations of the various types in Arthur's classification. The Gritsenko lifts are precisely those of Saito–Kurokawa type, while those of one-dimensional type do not contribute to the cuspidal spectrum. Since abelian surfaces with $\mathrm{End}_{\mathbf{Q}} A = \mathbf{Z}$ should correspond to automorphic representations of general type (as their corresponding Galois representations are irreducible), we see that it is implicit in the statement of the conjecture that there are no paramodular eigenforms (at least with rational Hecke eigenvalues) of Yoshida, Soudry, or Howe–Piatetski-Shapiro type.

This is indeed the case, as is proved in [Sch18, Sch20]. The case of Yoshida type is [Sch18, Lem. 2.5]; in this case, the parameters are unstable, and the corresponding packet of representations does not satisfy the required sign condition. Indeed, at each finite place, the condition that the representation admits a paramodular vector forces the sign to be trivial, whereas the condition of being the holomorphic limit of discrete series at infinity gives a non-trivial sign. Note that the analogous argument would fail for totally real fields of even degree.

The cases of Soudry and Howe–Piatetski-Shapiro type are [Sch20, Prop. 5.1]. In these cases the obstructions to the existence of paramodular vectors are at finite places; it turns out that at the places where these representations are ramified, there are no paramodular vectors. In these cases the representations are parameterized by certain Hecke characters, and the fact that the representations are ramified at some finite place comes from the fact that any Hecke character must be ramified. Accordingly, the analogous argument could fail for totally real fields of class number greater than 1.

It follows from this discussion that the original paramodular conjecture is equivalent to the claim that there is a bijection between isogeny classes of abelian surfaces $A/\mathbf{Q}$ with $\mathrm{End}_{\mathbf{Q}} A = \mathbf{Z}$, and cuspidal automorphic representations Π of $\mathrm{GL}_4(\mathbf{A}_{\mathbf{Q}})$ of symplectic type with multiplier $|\cdot|^2$, whose infinity type is the one corresponding to the L-parameter $\phi_{2;1,0}$, and whose Hecke eigenvalues are all rational. In one direction, given A, the existence of Π is certainly predicted by the Fontaine–Mazur–Langlands conjecture, the rationality of its Hecke eigenvalues following from strong multiplicity one. We now explore the converse direction.

Lemma 10.3.2. *Let F be a totally real field. Assume the Fontaine–Mazur Conjecture, the Standard Conjectures, the Hodge Conjecture, and that the Galois representations associated to any cuspidal automorphic representation Π for $\mathrm{GL}_4(\mathbf{A}_F)$ whose infinity type for each $v|\infty$ corresponds to the L-parameter $\phi_{2;1,0}$ form an irreducible weakly compatible system. Let Π be such a representation with the properties that its Hecke eigenvalues are rational, and that Π is of symplectic type with multiplier $|\cdot|^2$. Then, associated to Π, there exists a corresponding motive A/F such that either:*

(1) A/F is an abelian surface.
(2) A/F is an abelian fourfold with endomorphisms over F by an order in a quaternion algebra $D/\mathbf{Q}$.

Moreover, if A/F is an abelian fourfold with $\mathrm{End}_F(A) \otimes \mathbf{Q} = \mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q}$ an indefinite quaternion algebra $D/\mathbf{Q}$, and one assumes only standard automorphy conjectures, then there exists a corresponding Π of symplectic type with rational eigenvalues and multiplier $|\cdot|^2$.

One might reasonably (following Serre [DR73, §0.7, p. DeRa-13]) call an abelian fourfold A with endomorphisms by an order in a quaternion algebra $D/\mathbf{Q}$ a *fake* (or *false*) *abelian surface* (*fausse surface abélienne*).

Sketch of proof. (For a more detailed proof of a closely related result, see [PVZ16, Thm. 3.1].) One first obtains from Π a rank 4 symplectic weakly compatible irreducible family of p-adic Galois representations

$$\mathcal{R} = (\mathbf{Q}, S, \{Q_v(X)\}, \{r_p\}, \{H_\tau\})$$

with $H_\tau = (0,0,1,1)$ for all $\tau|\infty$, and such that

$$r_p : G_F \to \mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$$

has inverse cyclotomic similitude character. The Fontaine–Mazur conjecture implies that $\mathcal{R}$ arises from a pure irreducible motive M over F with coefficients in $\mathbf{Q}$ (we also now assume the standard conjectures [Kle94]). Concretely, this means that M is irreducible and that for each prime p, the p-adic étale realization of M, $H^i_{et}(M, \mathbf{Q}_p) \otimes_{\mathbf{Q}_p} \overline{\mathbf{Q}}_p$ contains r_p. By the Brauer–Nesbitt theorem, all the twists of

r_p by automorphisms of the coefficient field $\overline{\mathbf{Q}}_p$ are isomorphic to r_p. Therefore, if we assume the Tate conjecture, we deduce that $H^i_{et}(M, \mathbf{Q}_p) \otimes_{\mathbf{Q}_p} \overline{\mathbf{Q}}_p$ is a sum of copies of r_p. The rank of M is therefore $4d$ for some d. Let $\mathrm{End}_{\mathbf{Q}}(M) \otimes \mathbf{Q} = D$. Since M is simple, D is a division algebra. The centre of D is a number field E. We claim that $E = \mathbf{Q}$. It suffices to show that, for all p, the centre of $D \otimes_{\mathbf{Q}} \overline{\mathbf{Q}}_p$ is $\overline{\mathbf{Q}}_p$. By the Tate conjecture, however, we can determine $D \otimes \overline{\mathbf{Q}}_p$ from the endomorphisms of the p-adic étale realization of M, which is isomorphic to a direct sum of d copies of r_p. It follows that $\mathrm{End}_{\overline{\mathbf{Q}}_p[G_F]}(r_p^d)$ is a matrix algebra over $\overline{\mathbf{Q}}_p$, and thus has centre $\overline{\mathbf{Q}}_p$. Hence $E = \mathbf{Q}$.

Let us fix an embedding $F \to \mathbf{C}$. The Hodge realization of M is a polarized Hodge structure of weight one, which gives a polarized torus, and thus (by Riemann) an abelian variety B over $\mathbf{C}$. Since M is defined over F, we deduce that, for any automorphism σ of $\mathbf{C}$ over F, B^σ is isogenous to B. Let $2d$ be the dimension of B and r be the degree of its polarization. Let $\mathcal{A}_{2d,r}$ be the coarse moduli space of abelian varieties of dimension $2d$ with a degree r polarization. This is a scheme of finite type over $\mathbf{Q}$. We base change it to F. Let $[B]$ be the point on $\mathcal{A}_{2d,r}$ associated to B. The set of points on $\mathcal{A}_{2d,r}$ isogenous to $[B]$ is countable. On the other hand, if the residue field of $[B]$ in $\mathcal{A}_{2d,r}$ is transcendental over F, we deduce that its orbit under $\mathrm{Aut}(\mathbf{C}/F)$ is uncountable because there are uncountably many ways to embed a transcendental field of finite type over F into $\mathbf{C}$. It follows that $[B]$ is defined over a finite algebraic extension L of F and that B is defined over a finite algebraic extension L' of L. Moreover, $M|_{L'} = \mathfrak{h}^1(B)$. We now consider $C = \mathrm{Res}_{L'/F} B$, an abelian scheme over F. Looking at the p-adic étale realization C_p of C for a prime p, we find that $C_p = M_p \otimes \mathrm{Ind}_{L'/F}\mathbf{1}$ where M_p stands for the p-adic realization of M. We now let A be the simple factor of C whose p-adic étale realization contains r_p. Then $\mathcal{R}$ arises from $\mathfrak{h}^1(A)$. It follows from the Tate conjecture that $\mathfrak{h}^1(A)$ and the motive M we started with are in fact isomorphic.

Now taking into account that the centre of D is $\mathbf{Q}$, we deduce from the Albert classification (see [Mum08, Thm. 2, p. 201]) that A is one of the following three types:

(1) Type I: A/F is an abelian surface with $\mathrm{End}_{\mathbf{Q}}(A) = \mathbf{Z}$.
(2) Type II: A/F is an abelian fourfold with $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$, an indefinite quaternion algebra over $\mathbf{Q}$.
(3) Type III: A/F is an abelian fourfold with $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$, a definite quaternion algebra over $\mathbf{Q}$.

(Note that Type IV of the Albert classification cannot occur, because the centre $F = \mathbf{Q}$ of D is not a totally imaginary CM field.)

Suppose that $A/\mathbf{Q}$ is an abelian fourfold with $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = \mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q} = D$ for some indefinite quaternion algebra $D/\mathbf{Q}$ (and thus of Type II above). We now construct a suitable compatible system $\mathcal{R}$, which (by standard automorphy conjectures) will give rise to a suitable Π. The Mumford–Tate conjecture is known for the varieties of type II [Chi92, Chi90], and the semisimple part $\mathfrak{h}$ of the Lie algebra of the Mumford–Tate group of A is (for almost all p) $\mathfrak{sp}_4$. Let p be any prime which splits D. Then $H^1(A, \mathbf{Q}_p)$ has an action of $D \otimes \mathbf{Q}_p = M_2(\mathbf{Q}_p)$. In particular, it decomposes as $V_p \oplus V_p$ for an irreducible 4-dimensional representation V_p whose monodromy group is contained in $\mathrm{GSp}_4(\mathbf{Q}_p)$.

If $Q_v(T)$ denotes the degree 8 polynomial in $\mathbf{Z}[T]$ coming from the characteristic polynomial of Frobenius at v, then every root of $Q_v(T)$ has even multiplicity, and thus $Q_v(T) = P_v(T)^2$ for a degree 4 polynomial $P_v(T) \in \mathbf{Z}[T]$, which will be the characteristic polynomial of Frobenius at v on V_p. By the Weil conjectures, the roots of $Q_v(T)$ obey the usual symmetry associated to a weight one motive, and so the same is true for $P_v(T)$. This implies that $V_p^\vee \simeq V_p \otimes \varepsilon^{-1}$. Since the Galois representation has big image in V_p, any isomorphism $V_p \simeq V_p \otimes \chi$ forces χ to be trivial, and thus from the identification $V_p^\vee \simeq V_p \otimes \varepsilon^{-1}$ above we deduce that the similitude character is inverse cyclotomic. In particular, by standard automorphy conjectures, V will be associated with a Π as in the theorem. By strong multiplicity one [JS81], the rationality of Hecke eigenvalues at almost all primes (in particular primes of good reduction) forces rationality at all primes. □

10.4. **Examples and counterexamples.** In this section, we give some examples of abelian fourfolds A/F with $\mathrm{End}_F(A) \otimes \mathbf{Q} = D$ for a quaternion algebra $D/\mathbf{Q}$. In Lemma 10.4.4, we prove the existence of such A which also satisfy $\mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q} = D$ for some indefinite $D/\mathbf{Q}$. But first, we construct abelian fourfolds $A/\mathbf{Q}$ with $\mathrm{End}_F(A) \otimes \mathbf{Q} = D$, and such that (under standard conjectures, and even unconditionally in some cases), they correspond to a Π as above which comes from a paramodular eigenform with rational Hecke eigenvalues, and thus contradict the original paramodular conjecture.

10.4.1. *Abelian fourfolds of type III.* We expect that case (3) considered in the proof of Lemma 10.3.2 cannot occur. While we do not show that here, we instead discuss a minor subtlety which occurs when trying to construct examples of this kind.

Let $E/\mathbf{Q}$ be an elliptic curve (say without complex multiplication). Let $F/\mathbf{Q}$ be (say) a totally real field with $\mathrm{Gal}(F/\mathbf{Q}) = Q$, the quaternion group of order 8. The group Q has an irreducible representation $W/\mathbf{Q}$ of dimension 4, which contains a stable integral lattice $\Lambda \subset W$. Note that, for any prime p, there is a decomposition $W \otimes \overline{\mathbf{Q}}_p = V \oplus V$ for an irreducible 2-dimensional faithful representation V of Q. Now let us define:

$$A = E^4 = E \otimes_{\mathbf{Z}} \Lambda.$$

We find that A is simple over $\mathbf{Q}$, and $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = \mathrm{End}_{\mathbf{Q}}(W) = D$, where $D/\mathbf{Q}$ is the Hamilton quaternions. The corresponding compatible system $\mathcal{R}$ arises from Galois representations

$$\rho_p := r_{E,p} \otimes V : G_{\mathbf{Q}} \to \mathrm{GL}_4(\overline{\mathbf{Q}}_p).$$

For $p \neq 2$, the image of this Galois representation lies in $\mathrm{GL}_4(\mathbf{Q}_p)$. On the other hand, the possible symplectic forms associated to ρ_p are the one dimensional summands of

$$\wedge^2 \rho = (\mathrm{Sym}^2 \rho_{E,p} \otimes \det(V)) \oplus (\det(\rho_{E,p}) \otimes \mathrm{Sym}^2(V)).$$

Since $\det \rho_{E,p} = \varepsilon^{-1}$, we obtain symplectic representations with inverse cyclotomic similitude character if and only if $\mathrm{Sym}^2(V)$ contains the trivial representation. But $\det(V) = 1$ for the faithful complex 2-dimensional representation of Q, so $\mathrm{Sym}^2(V)$ is the direct sum of the three non-trivial quadratic characters of Q, and these compatible families do not have the required form.

More generally, suppose that $A/\mathbf{Q}$ is an abelian fourfold with $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$ for some definite quaternion algebra D. The corresponding Shimura curves X_D

parametrizing such objects lie in the exceptional class of Shimura varieties with the property that there is a strict containment $D \subsetneq \mathrm{End}_{\mathbf{C}}(A)$ for all complex points A of X_D (see [BL04, §9.9]). Since $D/\mathbf{Q}$ is definite, the semisimple part $\mathfrak{h}$ of the Lie algebra of the Mumford–Tate group should (for almost all p) be contained in $\mathfrak{so}_4 = \mathfrak{sl}_2 \times \mathfrak{sl}_2$ rather than $\mathfrak{sp}_4$ (by [MZ95, §6.1]). This forces the representations V_p to decompose (as $\overline{\mathbf{Q}}_p$-representations) as the tensor product of a representation coming from a modular form with an Artin representation. We expect it should be possible to make a careful case by case analysis to rule out this case occurring, but we have not attempted to do this.

10.4.2. *Abelian fourfolds of type II.* One can produce examples of abelian fourfolds with endomorphisms by an order in an indefinite quaternion algebra $D/\mathbf{Q}$ by taking the tensor product of a 2-dimensional representation with an Artin representation. Let $B/\mathbf{Q}$ be an abelian surface of GL(2)-type with endomorphisms by an order in a quaternion algebra D which are defined over a quadratic extension $K/\mathbf{Q}$, and then take $L \subset V$ to be a lattice in a 2-dimensional dihedral representation V over $\mathbf{Q}$ which is induced from a quadratic character χ of K which does not extend to $\mathbf{Q}$ (so the action of $G_{\mathbf{Q}}$ on V is through a dihedral group of order 8). Then one can take $A = B \otimes_{\mathbf{Z}} L$, which may be identified with the restriction of scalars of the quadratic twist $B \otimes \chi$ of B from K to $\mathbf{Q}$.

The action of an order of D on B and $B \otimes \chi$ over K extends to an action of this order of D on A. We obtain a compatible system $\mathcal{R}$ of Galois representations

$$\rho_p := r_{B,p} \otimes V : G_{\mathbf{Q}} \to \mathrm{GL}_4(\overline{\mathbf{Q}}_p).$$

Because V is induced from G_K, it follows that the characteristic polynomials of Frobenius of this representation all have coefficients in $\mathbf{Q}$. It now suffices to show that ρ_p preserves a symplectic form with inverse cyclotomic similitude character. The argument proceeds exactly as in §10.4.1, except now we have the isomorphism $\wedge^2 V \simeq \chi_{K/\mathbf{Q}}$. In particular, the trivial character is a summand of $\mathrm{Sym}^2 V$, and thus ρ_p preserves a symplectic form with similitude character ε^{-1}. The representations ρ_p do not arise from abelian surfaces over $\mathbf{Q}$, since that would contradict the Tate conjecture for abelian varieties [Fal83]. Moreover, they are easily seen to be modular. Hence these give counterexamples to the original paramodular conjecture.

For an explicit example, one could take B to be the modular abelian surface which is a quotient of $J_0(243)$ with coefficient field $\mathbf{Q}(\sqrt{6})$ (see [Cre92, Table 3]), which is geometrically simple and obtains quaternionic multiplication over $\mathbf{Q}(\sqrt{-3})$. Then take any non-Galois invariant quadratic character χ of $\mathbf{Q}(\sqrt{-3})$, and let $A = \mathrm{Res}_{K/\mathbf{Q}}(B \otimes \chi)$.

In light of Lemma 10.3.2, Brumer and Kramer have formulated the following natural modification of the original paramodular conjecture (see [BK19]):

Conjecture 10.4.3 (Paramodular Conjecture of Brumer–Kramer)**.** *Let $\mathcal{A}_N$ denote the set of isogeny classes of abelian surfaces $A/\mathbf{Q}$ with $\mathrm{End}_{\mathbf{Q}} A = \mathbf{Z}$ and conductor N, and $\mathcal{B}_N$ the set of isogeny classes of fake abelian surfaces (QM abelian fourfolds) $B/\mathbf{Q}$ of conductor N^2 with $\mathrm{End}_{\mathbf{Q}} B$ an order in a non-split quaternion algebra $D/\mathbf{Q}$. Let $\mathcal{P}_N$ denote the set of holomorphic weight 2 paramodular forms f of level N up to nonzero scaling which have rational Hecke eigenvalues and lie in the orthogonal complement to the space of Gritsenko lifts. Then there is a bijection*

between the set $\mathcal{A}_N \cup \mathcal{B}_N$ *and* $\mathcal{P}_N$ *such that*

$$L(C,s) = L(f,s,\mathrm{spin}) \text{ if } C \in \mathcal{A}_N \text{ and } L(C,s) = L(f,s,\mathrm{spin})^2 \text{ if } C \in \mathcal{B}_N.$$

We conclude with some remarks on the possible existence of abelian fourfolds which satisfy case (2) of Lemma 10.3.2 and additionally have no further endomorphisms over $\mathbf{C}$ (such varieties will necessarily be geometrically simple).

Expected Lemma 10.4.4. *There exists a totally real field F, an indefinite quaternion algebra D, and an abelian fourfold A/F with* $\mathrm{End}_F(A) \otimes \mathbf{Q} = \mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q} = D$.

Sketch. Let $D = \left(\frac{-1,3}{\mathbf{Q}}\right)$ be the unique quaternion algebra over $\mathbf{Q}$ ramified at (exactly) 2 and 3. Let $\mathcal{O}_D$ denote the maximal order in D. In the standard way, one may also write down an involution $\dagger$ obtained by conjugating the standard involution so that $\mathrm{tr}_{D/\mathbf{Q}}(xx^\dagger)$ is positive definite, and write down a non-degenerate alternating form ψ on $(\mathcal{O}_D^2) \otimes \mathbf{Q}$ which satisfies various compatibilities with $\dagger$. Associated to $\mathcal{O}_D$ in the usual way is a Shimura stack (of level one) X parametrizing tuples (A, λ, ι) where A is an S-abelian fourfold over S, λ is a principal polarization over S, and $\iota : \mathcal{O}_D \to \mathrm{End}(A)$ is an injective homomorphism such that the Rosati involution induced by λ restricts to $\dagger$ and such that ψ is compatible with the polarization on homology as an $\mathcal{O}_D$-module. X is a smooth Deligne–Mumford stack over $\mathbf{Q}$ with a single geometric component (cf. [KR99]). The complex points $X(\mathbf{C})$ are uniformized by the Siegel upper half space of dimension 3, and the generic point of X over $\mathbf{C}$ has endomorphisms precisely by $\mathcal{O}_D$ (see §9.9 of [BL04]).

By [Mil79], there exists an abelian surface $B/\mathbf{C}$ with $\mathrm{End}_{\mathbf{C}}(B) \otimes \mathbf{Q} = D$, that the Rosati involution on $\mathrm{End}\, B$ is $x \mapsto x^\dagger$, and such that the restriction of scalars of B from $\mathbf{C}$ to $\mathbf{R}$ gives a point in $X(\mathbf{R})$. (Another way to view this is to consider X as a GSpin Shimura variety associated to a 5 dimensional quadratic space as in [KR99], and then signature $(1,2)$ subspaces will give Shimura curve subvarieties.)

This is not quite sufficient, however, to guarantee a point over $X(F)$ for a totally real field F with the correct endomorphisms, nor even a point over $X(\overline{\mathbf{Q}})$, since one has to remove from $X(\mathbf{C})$ a countable union of proper Shimura subvarieties, which might *a priori* exhaust the $\overline{\mathbf{Q}}$-points of X. Moreover, due to the stackiness of X, there are issues comparing fields of definition versus fields of moduli. We therefore employ a trick already used in the proof of Theorem 10.2.1. Namely, impose level structure by choosing a large prime $p > 3$ and fixing a surjective representation $\overline{\rho} : G_{\mathbf{Q}} \to \mathrm{GSp}_4(\mathbf{F}_p)$ with inverse cyclotomic similitude character. (Such representations are abundant — one source are the duals of the p-torsion of abelian surfaces over $\mathbf{Q}$.) Then X admits a geometrically connected cover $X(\overline{\rho})$ defined over $\mathbf{Q}$ with level structure corresponding to $A[p]^\vee = \overline{\rho} \oplus \overline{\rho}$, with a suitable choice of polarization and compatible action of $(\mathcal{O}_D \otimes \mathbf{Z}_p)/p = (M_2(\mathbf{Z}_p))/p = M_2(\mathbf{F}_p)$.

The variety $X(\overline{\rho})$ is a fine moduli space which is now a smooth variety over $\mathbf{Q}$ with real points, since the point $\mathrm{Res}_{\mathbf{C}/\mathbf{R}}(B)$ considered above has the appropriate level structure over $\mathbf{R}$. Employing the theorem of Moret-Bailly [MB89], we may deduce the existence of a totally real field F and a corresponding abelian variety A/F such that F is disjoint from the splitting field of $\overline{\rho}$. Because $X(\overline{\rho})$ is a fine moduli space, the variety A has endomorphisms by $\mathcal{O}_D$ over F. It now suffices to show that it has no further endomorphisms over $\mathbf{C}$. The dual of the Tate module of A

decomposes as a Galois representation as $\rho \oplus \rho$ where $\rho : G_F \to \mathrm{GSp}_4(\mathbf{Z}_p)$ is a lift of $\overline{\rho}$. Since $p \geq 5$, the assumption that $\overline{\rho}$ has surjective image implies that ρ also has surjective image. However, if A admitted extra endomorphisms over any extension of F, then the image of ρ restricted to some open subgroup would lie inside a proper algebraic subgroup of $\mathrm{GSp}_4(\mathbf{Q}_p)$, contradicting the fact that image contains an open subgroup of $\mathrm{GSp}_4(\mathbf{Z}_p)$. □

It would be interesting to know whether (for suitable choices) these varieties have points over $\mathbf{Q}$ which correspond to $A/\mathbf{Q}$ with $\mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q} = D$, but this is not so easy to determine by pure thought. However, the specific X chosen above (ramified at only 2 and 3) is possibly the most likely choice to be rational, since it corresponds to the indefinite quaternion algebra $D/\mathbf{Q}$ of smallest discriminant. The construction of Nori in §9.4.4 suggests that, for this D, the moduli space is at least geometrically rational. Note, however, that there will be field of moduli issues when one works at level one, so even the rationality of this space over $\mathbf{Q}$ does not imply the existence of such A.

10.4.5. *Cremona's Question.* We finally consider two 2-dimensional irreducible compatible systems $\mathcal{S}$ of representations of G_K for some quadratic extension $K/\mathbf{Q}$, with inverse cyclotomic determinant, Hodge–Tate weights $(0, 1)$, and coefficients in $\mathbf{Q}$. Note that, for such a family $\mathcal{S}$, there is a corresponding family $\mathcal{R} = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}}} \mathcal{S}$ of 4-dimensional symplectic representations with inverse cyclotomic similitude character. An argument very similar to (but easier than) Lemma 10.3.2 shows that (assuming all conjectures) either $\mathcal{S}$ comes from an elliptic curve, or it arises from a so-called fake elliptic curve, namely, an abelian surface B/K with $\mathrm{End}_K(B) \otimes \mathbf{Q} = D$ for some indefinite quaternion algebra D. The latter can exist only when K is an imaginary quadratic field. Conjecturally, such compatible systems are in bijection with cuspidal cohomological $\boldsymbol{\pi}$ for $\mathrm{GL}(2)/K$ with trivial central character and Hecke eigenvalues in $\mathbf{Q}$.

One source of such B/K is to take abelian surfaces over $\mathbf{Q}$ of $\mathrm{GL}(2)$-type which acquire quaternionic multiplication over $K/\mathbf{Q}$. Assuming the Hodge conjecture and the standard conjectures, it follows that [Cre92, Question 1′] (cf. [Cre84, Conjecture, p.278]) is equivalent to asking that *all* fake elliptic curves B over K descend to $\mathbf{Q}$ after twisting by some quadratic character (equivalently, B is isogenous to a twist of B^σ for the non-trivial element $\sigma \in \mathrm{Gal}(K/\mathbf{Q})$). We call such B *non-autochthonous* because it implies that the corresponding conjectural $\boldsymbol{\pi}$ arises via functoriality from a smaller rank group (cf. footnote 2 of [AGM20]). In this section, we show that the answer to this question is false, namely, we construct autochthonous fake elliptic curves B/K. If one takes the restriction of scalars $A = \mathrm{Res}_{K/\mathbf{Q}}(B)$ of such surfaces, then the fourfolds A give rise to further examples in opposition to the original paramodular conjecture.

We continue to let D be the quaternion algebra ramified at precisely 2 and 3. The Shimura curve giving rise to fake elliptic curves with endomorphisms by a maximal order in $\mathcal{O}_D$ has genus zero, and is well-known (see for example [BG08, Thm. 11]) to be isomorphic over $\mathbf{Q}$ to:

$$X^2 + Y^2 + 3Z^2 = 0.$$

Moreover, more usefully for our purposes, Baba and Granath in [BG08] give explicit models for genus two curves with endomorphisms by $\mathcal{O}_D$. For a certain parameter j,

they write down a model ([BG08, Thm. 15]) of a genus two curve C over $\mathbf{Q}(\sqrt{-6j})$ such that its endomorphisms are all defined (by [BG08, Prop. 19]) over $K := \mathbf{Q}(\sqrt{-6j}, \sqrt{j}, \sqrt{-27(j+16)})$. With a view to choosing $K = \mathbf{Q}(\sqrt{-6})$, we let

$$Z = 3\sqrt{j}, \qquad X = \sqrt{-(27j+16)}, \qquad Y = 4,$$

and look for solutions to the equation above with $X, Z \in \mathbf{Q}(\sqrt{-6})$. One such solution is given by $j = -32/27$, but the corresponding surface is not autochthonous. Thus, we parametrize the conic and choose a random such point. Without making too much effort to optimize the height of j, one finds that

$$j = \frac{4(1+2\sqrt{-6})}{27} \tag{10.4.5}$$

is a suitable point.

Lemma 10.4.6. *Let C be the following genus two curve:*

$$y^2 = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6,$$

where a_i are given by the following table, where $\eta = 1 - \sqrt{-6}$:

i	a_i
0	$2^{10} \cdot \eta^6 \cdot (\sqrt{-6} - 4)$
1	$2^{10} \cdot \eta^6 \cdot 3$
2	$2^7 \cdot \eta^4 \cdot (9\sqrt{-6} + 24)$
3	$-2^8 \cdot \eta^4$
4	$2^4 \cdot \eta^2 \cdot (-9\sqrt{-6} + 60)$
5	$2^4 \cdot \eta \cdot 3$
6	$-2 \cdot \sqrt{-6}$

The sextic has discriminant $2^{90} \cdot 3^6 \cdot (1-\sqrt{-6})^{30} \cdot (2-\sqrt{-6})^6$. Let $B = \mathrm{Jac}(C)/\mathbf{Q}(\sqrt{-6})$, and let $A = \mathrm{Res}_{\mathbf{Q}(\sqrt{-6})/\mathbf{Q}}(B)$. Then $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$ and $\mathrm{End}_{\mathbf{C}}(A) \otimes \mathbf{Q} = M_2(D)$, where $D/\mathbf{Q}$ is the quaternion algebra ramified at precisely 2 and 3. Then B is autochthonous, and B gives rise to an irreducible 2-dimensional compatible system of Galois representations $\mathcal{S}$ of $G_{\mathbf{Q}}$ with Hodge–Tate weights $(0,1)$ and inverse cyclotomic determinant, and A gives rise to a 4-dimensional compatible system $\mathcal{R}$ of 4-dimensional p-adic Galois representations of $G_{\mathbf{Q}}$ with coefficients in $\mathbf{Q}$ unramified outside of $\{2,3,5,7,p\}$, each of which is absolutely irreducible and symplectic with inverse cyclotomic multiplier.

Proof. Let $K = \mathbf{Q}(\sqrt{-6})$. The curve C is the specialization of the curve in [BG08, Thm. 15] to the parameter j as in equation 10.4.5, and

$$s = \sqrt{-6j} = \frac{-2\sqrt{-6}+4}{3}, \qquad t = -2(27j+16) = -16\sqrt{-6} - 40.$$

By [BG08, Prop. 19], we deduce that the endomorphisms of C are defined over the field $K(\sqrt{j}, \sqrt{-(27j+16)}) = K$. We now show that B is autochthonous. Let $\mathfrak{p} = (11, \sqrt{-6} - 4)$ and $\mathfrak{q} = (11, \sqrt{-6} + 4)$. Then the curves $X_1 = C(\mathcal{O}_K/\mathfrak{p})$ and $X_2 = C(\mathcal{O}_K/\mathfrak{q})$ over $\mathbf{F}_{11}$ are given explicitly as follows:

$$X_1 : y^2 = 3x^6 + 3x^5 + 2x^4 + 10x^3 + 8x^2 + 9x,$$

$$X_2 : y^2 = 8x^6 + x^5 + 10x^4 + 6x^3 + 3x^2 + 4x + 4.$$

We compute the zeta functions using `magma` (see [BCGP21]) to be as follows:

$$Z(X_1, s) = \frac{(1 - 2p^{-s} + p^{1-2s})^2}{(1 - p^{-s})(1 - p^{1-s})}, \quad Z(X_2, s) = \frac{(1 - p^{-s} + p^{1-2s})^2}{(1 - p^{-s})(1 - p^{1-s})}.$$

If B were autochthonous, then in particular the zeta functions of X_1, X_2 would differ by a twist by a finite order character, but this is impossible since $2 \neq \pm 1$.

If $p > 3$, then p splits in D, so there are Galois representations

$$r_p : G_K \to \mathrm{GL}_2(\mathbf{Q}_p)$$

with $V_p(B)^\vee \cong r_p \oplus r_p$. For all p, there also exist corresponding representations $\widetilde{r}_p : G_K \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ such that, for $p > 3$, the representation $\widetilde{r}_p$ is the representation obtained from r_p by extending scalars. We now prove that $\mathrm{End}_{\mathbf{C}}(B) \otimes \mathbf{Q} = D$. If this were not true, then B would geometrically have to be isogenous to $E \times E$ for some elliptic curve E with complex multiplication. This implies that B itself has complex multiplication over $\mathbf{C}$, which implies that the representations r_p are potentially reducible. But as the representations r_p have distinct Hodge–Tate weights, if they become reducible they do so over a quadratic extension. This quadratic extension L/K must be ramified only at primes of bad reduction of B, and for $\mathfrak{p}$ which are inert in L/K, one must have $a_{\mathfrak{p}} = 0$. But this can be ruled out by computation (the only prime $\mathfrak{p}$ of norm less than 1000 with $a_{\mathfrak{p}} = 0$ has norm 97).

Hence $\mathrm{End}_{\mathbf{C}}(B) \otimes \mathbf{Q} = D$ and $\mathrm{End}_{\mathbf{Q}}(A) \otimes \mathbf{Q} = D$, where $A = \mathrm{Res}_{K/\mathbf{Q}}(B)$. It also follows that the representations r_p and $\widetilde{r}_p$ have inverse cyclotomic determinant (as otherwise they would be isomorphic to their twists by a finite order character, and thus potentially reducible). Moreover, with $\rho_p := \mathrm{Ind}_K^{\mathbf{Q}} \widetilde{r}_p$, one has

$$\wedge^2 \rho_p = \mathrm{As}(\widetilde{r}_p) \oplus \varepsilon^{-1} \oplus \varepsilon^{-1} \cdot \eta_{K/\mathbf{Q}}, \quad \mathrm{Sym}^2 \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}}} \rho_p = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}}} \mathrm{Sym}^2 \widetilde{r}_p,$$

and thus ρ_p is absolutely irreducible and can be chosen to have image in $\mathrm{GSp}_4(\overline{\mathbf{Q}}_p)$ with inverse cyclotomic similitude character. Finally, the characteristic polynomials of Frobenius will, by construction, be degree 4 polynomials with coefficients in $\mathbf{Q}$. □

Since one expects the compatible system $\mathcal{S}$ to be modular (it is certainly potentially modular, by [ACC+18]), it follows that Cremona's question [Cre92, Question 1′]) is incompatible with standard modularity conjectures. (Similarly, the modularity of $\mathcal{R} = \mathrm{Ind}_{G_K}^{G_{\mathbf{Q}}} \mathcal{S}$ is incompatible with the original paramodular conjecture, although we have already shown the latter to be false.) Of course, from the discussion above, there are natural modifications that one could make to Cremona's question (along the lines of Conjecture 10.4.3) — namely, to include all fake elliptic curves over K, autochthonous or otherwise.

One can presumably show that the 2-dimensional G_K-representations $\widetilde{r}_p$ over the field $K = \mathbf{Q}(\sqrt{-6})$ arising from $B = \mathrm{Jac}(C)$ are modular for $\mathrm{GL}(2)/K$. As in the proof of Lemma 10.3.1, we would then obtain a cuspidal cohomological automorphic representation $\boldsymbol{\pi}$ for $\mathrm{GL}(2)/K$ with trivial central character and rational eigenvalues. Since $\boldsymbol{\pi}$ does not arise (up to twist) from base change, this would answer in the negative [Cre92, Question 1′], because the existence of a corresponding elliptic curve E/K would be incompatible with the existence of B by Faltings' isogeny theorem [Fal83].

The modularity of the representations $\widetilde{r}_p$ can in principle be established using the Faltings–Serre method (cf. [BDPcS15]). Possibly some computational advantage

would be gained by replacing C with a curve obtained from a more careful choice of generic point on the Shimura curve (in order to work at a manageable level). As it turns out, Ciaran Schembri [Sch19] has independently found examples of autochthonous fake elliptic curves which he has verified are modular.

References

[AC89] James Arthur and Laurent Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies, vol. 120, Princeton University Press, Princeton, NJ, 1989.

[ACC+18] Patrick Allen, Frank Calegari, Ana Caraiani, Toby Gee, David Helm, Bao Le Hung, James Newton, Peter Scholze, Richard Taylor, and Jack Thorne, *Potential automorphy over CM fields*, preprint, 2018.

[AGM20] Avner Ash, Paul E. Gunnells, and Mark McConnell, *Cohomology with twisted one-dimensional coefficients for congruence subgroups of* $\mathrm{SL}_4(\mathbf{Z})$ *and Galois representations*, J. Algebra **553** (2020), 211–247.

[AIP15] Fabrizio Andreatta, Adrian Iovita, and Vincent Pilloni, *p-adic families of Siegel modular cuspforms*, Ann. of Math. (2) **181** (2015), no. 2, 623–697.

[All16] Patrick B. Allen, *Deformations of polarized automorphic Galois representations and adjoint Selmer groups*, Duke Math. J. **165** (2016), no. 13, 2407–2460.

[AMRT10] Avner Ash, David Mumford, Michael Rapoport, and Yung-Sheng Tai, *Smooth compactifications of locally symmetric varieties*, second ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2010, With the collaboration of Peter Scholze.

[Art04] James Arthur, *Automorphic representations of* GSp(4), Contributions to automorphic forms, geometry, and number theory, Johns Hopkins Univ. Press, Baltimore, MD, 2004, pp. 65–81.

[Art13] ———, *The endoscopic classification of representations*, American Mathematical Society Colloquium Publications, vol. 61, American Mathematical Society, Providence, RI, 2013, Orthogonal and symplectic groups.

[AS08] Avner Ash and Glenn Stevens, *p-adic deformations of arithmetic cohomology*, preprint, 2008.

[AT09] Emil Artin and John Tate, *Class field theory*, AMS Chelsea Publishing, Providence, RI, 2009, Reprinted with corrections from the 1967 original.

[BÖ0] Gebhard Böckle, *Demuškin groups with group actions and applications to deformations of Galois representations*, Compositio Math. **121** (2000), no. 2, 109–154.

[Bak46] Henry F. Baker, *A Locus with* 25920 *Linear Self-Transformations*, Cambridge Tracts in Mathematics and Mathematical Physics, no. 39, Cambridge, at the University Press; New York, The Macmillan Company, 1946.

[Bal12] Sundeep Balaji, *G-valued potentially semi-stable deformation rings*, ProQuest LLC, Ann Arbor, MI, 2012, Thesis (Ph.D.)–The University of Chicago.

[BC11] Joël Bellaïche and Gaëtan Chenevier, *The sign of Galois representations attached to automorphic forms for unitary groups*, Compos. Math. **147** (2011), no. 5, 1337–1352.

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939 (electronic).

[BCGP21] George Boxer, Frank Calegari, Toby Gee, and Vincent Pilloni, *Auxiliary magma files*, `https://github.com/fcale75/abeliansurfacesmagmafiles`, 2021.

[BCP97] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993).

[BDPcS15] Tobias Berger, Lassina Dembélé, Ariel Pacetti, and Mehmet Haluk ¸Sengün, *Theta lifts of Bianchi modular forms and applications to paramodularity*, J. Lond. Math. Soc. (2) **92** (2015), no. 2, 353–370.

[BG08] Srinath Baba and Håkan Granath, *Genus 2 curves with quaternionic multiplication*, Canad. J. Math. **60** (2008), no. 4, 734–757.

[BG14] Kevin Buzzard and Toby Gee, *The conjectural connections between automorphic representations and Galois representations*, Automorphic forms and Galois representations. Vol. 1, London Math. Soc. Lecture Note Ser., vol. 414, Cambridge Univ. Press, Cambridge, 2014, pp. 135–187.

[BG19] Rebecca Bellovin and Toby Gee, *G-valued local deformation rings and global lifts*, Algebra Number Theory **13** (2019), no. 2, 333–378.

[BHR94] Don Blasius, Michael Harris, and Dinakar Ramakrishnan, *Coherent cohomology, limits of discrete series, and Galois conjugation*, Duke Math. J. **73** (1994), no. 3, 647–685.

[BJ15] Gebhard Böckle and Ann-Kristin Juschka, *Irreducibility of versal deformation rings in the (p, p)-case for 2-dimensional representations*, J. Algebra **444** (2015), 81–123.

[BK14] Armand Brumer and Kenneth Kramer, *Paramodular abelian varieties of odd conductor*, Trans. Amer. Math. Soc. **366** (2014), no. 5, 2463–2516.

[BK19] ———, *Corrigendum to "Paramodular abelian varieties of odd conductor"*, Trans. Amer. Math. Soc. **372** (2019), no. 3, 2251–2254.

[BK20] Tobias Berger and Krzysztof Klosin, *Deformations of Saito-Kurokawa type and the paramodular conjecture*, Amer. J. Math. **142** (2020), no. 6, 1821–1875, With and appendix by Chris Poor, Jerry Shurman, and David S. Yuen.

[BL04] Christina Birkenhake and Herbert Lange, *Complex abelian varieties*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 302, Springer-Verlag, Berlin, 2004.

[Bla06] Don Blasius, *Hilbert modular forms and the Ramanujan conjecture*, Noncommutative geometry and number theory, Aspects Math., E37, Vieweg, Wiesbaden, 2006, pp. 35–56.

[BLGG13] Thomas Barnet-Lamb, Toby Gee, and David Geraghty, *Congruences between Hilbert modular forms: constructing ordinary lifts, II*, Math. Res. Lett. **20** (2013), no. 1, 67–72.

[BLGGT14a] Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor, *Local-global compatibility for $l = p$, II*, Ann. Sci. Éc. Norm. Supér. (4) **47** (2014), no. 1, 165–179.

[BLGGT14b] ———, *Potential automorphy and change of weight*, Ann. of Math. (2) **179** (2014), no. 2, 501–609.

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

[BN18] Nils Bruin and Brett Nasserden, *Arithmetic aspects of the Burkhardt quartic threefold*, J. Lond. Math. Soc. (2) **98** (2018), no. 3, 536–556.

[Bol07] Michele Bolognesi, *On Weddle surfaces and their moduli*, Adv. Geom. **7** (2007), no. 1, 113–144.

[Box15] George Boxer, *Torsion in the Coherent Cohomology of Shimura Varieties and Galois Representations*, preprint, July 2015.

[BPP+19] Armand Brumer, Ariel Pacetti, Cris Poor, Gonzalo Tornaría, John Voight, and David S. Yuen, *On the paramodularity of typical abelian surfaces*, Algebra Number Theory **13** (2019), no. 5, 1145–1195.

[BPS16] Stéphane Bijakowski, Vincent Pilloni, and Benoît Stroh, *Classicité de formes modulaires surconvergentes*, Ann. of Math. (2) **183** (2016), no. 3, 975–1014.

[BPY16] Jeffery Breeding, II, Cris Poor, and David S. Yuen, *Computations of spaces of paramodular forms of general level*, J. Korean Math. Soc. **53** (2016), no. 3, 645–689.

[BR16] Riccardo Brasca and Giovanni Rosso, *Eigenvarieties for non-cuspidal modular forms over certain PEL Shimura varieties*, 2016.

[Bra47] Richard Brauer, *On Artin's L-series with general group characters*, Ann. of Math. (2) **48** (1947), 502–514.

[BT99] Kevin Buzzard and Richard Taylor, *Companion forms and weight one forms*, Ann. of Math. (2) **149** (1999), no. 3, 905–919.

[Bur91] Heinrich Burkhardt, *Untersuchungen aus dem Gebiete der hyperelliptischen Modulfunctionen*, Math. Ann. **38** (1891), no. 2, 161–224.

[Buz03] Kevin Buzzard, *Analytic continuation of overconvergent eigenforms*, J. Amer. Math. Soc. **16** (2003), no. 1, 29–55.

[Cai14] Bryden Cais, *On the U_p operator in characteristic p*, Math. Res. Lett. **21** (2014), no. 2, 255–260.

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

[Cal18] ———, *Non-minimal modularity lifting in weight one*, J. Reine Angew. Math. **740** (2018), 41–62.

[Car12] Ana Caraiani, *Local-global compatibility and the action of monodromy on nearby cycles*, Duke Math. J. **161** (2012), no. 12, 2311–2413.

[Car14] ———, *Monodromy and local-global compatibility for $l = p$*, Algebra Number Theory **8** (2014), no. 7, 1597–1646.

[Cas] William Casselman, *Introduction to admissible representations of p-adic groups*, Available at `https://www.math.ubc.ca/~cass/research/publications.html`.

[CC20] Frank Calegari and Shiva Chidambaram, *Rationality of twists of the Siegel modular variety of genus* 2 *and level* 3, https://arxiv.org/abs/2009.00194, 2020.

[CCG20] Frank Calegari, Shiva Chidambaram, and Alexandru Ghitza, *Some modular abelian surfaces*, Math. Comp. **89** (2020), no. 321, 387–394.

[CCN+85] John H. Conway, Robert T. Curtis, Simon P. Norton, Richard A. Parker, and Robert A. Wilson, *Atlas of finite groups*, Oxford University Press, Eynsham, 1985, Maximal subgroups and ordinary characters for simple groups, With computational assistance from J. G. Thackray.

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

[CG13] Frank Calegari and Toby Gee, *Irreducibility of automorphic Galois representations of $GL(n)$, n at most 5*, Ann. Inst. Fourier (Grenoble) **63** (2013), no. 5, 1881–1912.

[CG18] Frank Calegari and David Geraghty, *Modularity lifting beyond the Taylor–Wiles method*, Inventiones mathematicae **211** (2018), no. 1, 297–433.

[CG20] Frank Calegari and David Geraghty, *Minimal modularity lifting for nonregular symplectic representations*, Duke Math. J. **169** (2020), no. 5, 801–896, With an appendix by Calegari, Geraghty and Michael Harris.

[Che04] Gaëtan Chenevier, *Familles p-adiques de formes automorphes pour* GL_n, J. Reine Angew. Math. **570** (2004), 143–217.

[Chi90] Wên Chên Chi, *On the l-adic representations attached to some absolutely simple abelian varieties of type* II, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **37** (1990), no. 2, 467–484.

[Chi92] ———, *l-adic and λ-adic representations associated to abelian varieties defined over number fields*, Amer. J. Math. **114** (1992), no. 2, 315–353.

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

[CL10] Pierre-Henri Chaudouard and Gérard Laumon, *Le lemme fondamental pondéré. I. Constructions géométriques*, Compos. Math. **146** (2010), no. 6, 1416–1506.

[Clo90] Laurent Clozel, *Motifs et formes automorphes: applications du principe de fonctorialité*, Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988), Perspect. Math., vol. 10, Academic Press, Boston, MA, 1990, pp. 77–159.

[Cob06] Arthur B. Coble, *An Invariant Condition for Certain Automorphic Algebraic Forms*, Amer. J. Math. **28** (1906), no. 4, 333–366.

[Col97] Robert F. Coleman, *P -adic Banach spaces and families of modular forms*, Inventiones Mathematicae **127** (1997), 417–479.

[Cre84] John E. Cremona, *Hyperbolic tessellations, modular symbols, and elliptic curves over complex quadratic fields*, Compositio Math. **51** (1984), no. 3, 275–324.

[Cre92] ———, *Abelian varieties with extra twist, cusp forms, and elliptic curves over imaginary quadratic fields*, J. London Math. Soc. (2) **45** (1992), no. 3, 404–416.

[DDT97] Henri Darmon, Fred Diamond, and Richard Taylor, *Fermat's last theorem*, Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), Int. Press, Cambridge, MA, 1997, pp. 2–140.

[Del73] Pierre Deligne, *Les constantes des équations fonctionnelles des fonctions L*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp,

Antwerp, 1972), Springer, Berlin, 1973, pp. 501–597. Lecture Notes in Math., Vol. 349.

[Del79] ———, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 247–289.

[Dia97] Fred Diamond, *The Taylor–Wiles construction and multiplicity one*, Invent. Math. **128** (1997), no. 2, 379–391.

[dJ93] Aise J. de Jong, *The moduli spaces of principally polarized abelian varieties with $\Gamma_0(p)$-level structure*, J. Algebraic Geom. **2** (1993), no. 4, 667–688.

[DK00] William Duke and Emmanuel Kowalski, *A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations*, Invent. Math. **139** (2000), no. 1, 1–39, With an appendix by Dinakar Ramakrishnan.

[DK16] Lassina Dembélé and Abhinav Kumar, *Examples of abelian surfaces with everywhere good reduction*, Math. Ann. **364** (2016), no. 3-4, 1365–1392.

[DK17] Fred Diamond and Payman L. Kassaei, *Minimal weights of Hilbert modular forms in characteristic p*, Compos. Math. **153** (2017), no. 9, 1769–1778.

[DL08] Igor Dolgachev and David Lehavi, *On isogenous principally polarized abelian surfaces*, Curves and abelian varieties, Contemp. Math., vol. 465, Amer. Math. Soc., Providence, RI, 2008, pp. 51–69.

[DR73] Pierre Deligne and Michael Rapoport, *Les schémas de modules de courbes elliptiques*, Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 143–316. Lecture Notes in Math., Vol. 349.

[Edi92] Bas Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), no. 3, 563–594.

[EG14] Matthew Emerton and Toby Gee, *A geometric perspective on the Breuil-Mézard conjecture*, J. Inst. Math. Jussieu **13** (2014), no. 1, 183–223.

[Ell05] Jordan S. Ellenberg, *Serre's conjecture over $\mathbf{F}_9$*, Ann. of Math. (2) **161** (2005), no. 3, 1111–1142.

[ERX17] Matthew Emerton, Davide Reduzzi, and Liang Xiao, *Unramifiedness of Galois representations arising from Hilbert modular surfaces*, Forum Math. Sigma **5** (2017), e29, 70.

[Fal83] Gerd Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366.

[Far08] Laurent Fargues, *L'isomorphisme entre les tours de Lubin-Tate et de Drinfeld et applications cohomologiques*, L'isomorphisme entre les tours de Lubin-Tate et de Drinfeld, Progr. Math., vol. 262, Birkhäuser, Basel, 2008, pp. 1–325.

[Far10] ———, *La filtration de Harder-Narasimhan des schémas en groupes finis et plats*, J. Reine Angew. Math. **645** (2010), 1–39.

[Far11] ———, *La filtration canonique des points de torsion des groupes p-divisibles*, Ann. Sci. Éc. Norm. Supér. (4) **44** (2011), no. 6, 905–961, With collaboration of Yichao Tian.

[FC90] Gerd Faltings and Ching-Li Chai, *Degeneration of abelian varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 22, Springer-Verlag, Berlin, 1990, With an appendix by David Mumford.

[FKRS12] Francesc Fité, Kiran S. Kedlaya, Víctor Rotger, and Andrew V. Sutherland, *Sato-Tate distributions and Galois endomorphism modules in genus 2*, Compos. Math. **148** (2012), no. 5, 1390–1442.

[Fon94] Jean-Marc Fontaine, *Représentations l-adiques potentiellement semi-stables*, Astérisque (1994), no. 223, 321–347, Périodes p-adiques (Bures-sur-Yvette, 1988).

[FP21] Najmuddin Fakhruddin and Vincent Pilloni, *Hecke operators and the coherent cohomology of Shimura varieties*, Journal of the Institute of Mathematics of Jussieu (2021), 1–69.

[Fre83] Eberhard Freitag, *Siegelsche Modulfunktionen*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 254, Springer-Verlag, Berlin, 1983.

[Ger19] David Geraghty, *Modularity lifting theorems for ordinary Galois representations*, Math. Ann. **373** (2019), no. 3-4, 1341–1427.

[GG12] Toby Gee and David Geraghty, *Companion forms for unitary and symplectic groups*, Duke Math. J. **161** (2012), no. 2, 247–303.

[GHLS17] Toby Gee, Florian Herzig, Tong Liu, and David Savitt, *Potentially crystalline lifts of certain prescribed types*, Doc. Math. **22** (2017), 397–422.

[GJ72] Roger Godement and Hervé Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Mathematics, Vol. 260, Springer-Verlag, Berlin-New York, 1972.

[GK19] Wushi Goldring and Jean-Stefan Koskivirta, *Strata Hasse invariants, Hecke algebras and Galois representations*, Invent. Math. **217** (2019), no. 3, 887–984.

[GN20] Toby Gee and James Newton, *Patching and the completed homology of locally symmetric spaces*, Journal of the Institute of Mathematics of Jussieu (2020), 1–64.

[GPSR87] Stephen Gelbart, Ilya Piatetski-Shapiro, and Stephen Rallis, *Explicit constructions of automorphic L-functions*, Lecture Notes in Mathematics, vol. 1254, Springer-Verlag, Berlin, 1987.

[Gro90] Benedict H. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J. **61** (1990), no. 2, 445–517.

[Gro16] ———, *On the Langlands correspondence for symplectic motives*, Izv. Ross. Akad. Nauk Ser. Mat. **80** (2016), no. 4, 49–64.

[GRR72] Alexander Grothendieck, Michèle Raynaud, and Dock S. Rim, *Groupes de monodromie en géométrie algébrique. I*, Lecture Notes in Mathematics, Vol. 288, Springer-Verlag, Berlin-New York, 1972, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 I), Dirigé par A. Grothendieck. Avec la collaboration de Michèle Raynaud et Dock S. Rim.

[GT05] Alain Genestier and Jacques Tilouine, *Systèmes de Taylor-Wiles pour* GSp(4), Astérisque (2005), no. 302, 177–290, Formes automorphes. II. Le cas du groupe GSp(4).

[GT11a] Wee Teck Gan and Shuichiro Takeda, *The local Langlands conjecture for* GSp(4), Ann. of Math. (2) **173** (2011), no. 3, 1841–1882.

[GT11b] ———, *Theta correspondences for* GSp(4), Represent. Theory **15** (2011), 670–718.

[GT19] Toby Gee and Olivier Taïbi, *Arthur's multiplicity formula for* $\mathbf{GSp}_4$ *and restriction to* $\mathbf{Sp}_4$, J. Éc. polytech. Math. **6** (2019), 469–535.

[Har66] Robin Hartshorne, *Residues and duality*, Lecture notes of a seminar on the work of A. Grothendieck, given at Harvard 1963/64. With an appendix by Pierre Deligne. Lecture Notes in Mathematics, No. 20, Springer-Verlag, Berlin-New York, 1966.

[Har90a] Michael Harris, *Automorphic forms and the cohomology of vector bundles on Shimura varieties*, Automorphic forms, Shimura varieties, and L-functions, Vol. II (Ann Arbor, MI, 1988), Perspect. Math., vol. 11, Academic Press, Boston, MA, 1990, pp. 41–91.

[Har90b] ———, *Automorphic forms of $\overline{\partial}$-cohomology type as coherent cohomology classes*, J. Differential Geom. **32** (1990), no. 1, 1–63.

[Hec20] Erich Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen*, Math. Z. **6** (1920), no. 1-2, 11–51.

[Hen09] Guy Henniart, *Sur la fonctorialité, pour* GL(4)*, donnée par le carré extérieur*, Mosc. Math. J. **9** (2009), no. 1, 33–45, back matter.

[Hid04] Haruzo Hida, *p-adic automorphic forms on Shimura varieties*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2004.

[Hid09] ———, *Irreducibility of the Igusa tower*, Acta Math. Sin. (Engl. Ser.) **25** (2009), no. 1, 1–20.

[HKP10] Thomas J. Haines, Robert E. Kottwitz, and Amritanshu Prasad, *Iwahori-Hecke algebras*, J. Ramanujan Math. Soc. **25** (2010), no. 2, 113–145.

[HLTT16] Michael Harris, Kai-Wen Lan, Richard Taylor, and Jack Thorne, *On the rigid cohomology of certain Shimura varieties*, Res. Math. Sci. **3** (2016), 3:37.

[HS02] Klaus Hulek and G. K. Sankaran, *The geometry of Siegel modular varieties*, Higher dimensional birational geometry (Kyoto, 1997), Adv. Stud. Pure Math., vol. 35, Math. Soc. Japan, Tokyo, 2002, pp. 89–156.

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

[Hub96] Roland Huber, *étale cohomology of rigid analytic varieties and adic spaces*, Aspects of Mathematics, E30, Friedr. Vieweg & Sohn, Braunschweig, 1996.

[Hun96] Bruce Hunt, *The geometry of some special arithmetic quotients*, Lecture Notes in Mathematics, vol. 1637, Springer-Verlag, Berlin, 1996.

[HZ01] Michael Harris and Steven Zucker, *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, vi+116.

[Igu64] Jun-ichi Igusa, *On Siegel modular forms genus two. II*, Amer. J. Math. **86** (1964), 392–412.

[JLR12] Jennifer Johnson-Leung and Brooks Roberts, *Siegel modular forms of degree two attached to Hilbert modular forms*, J. Number Theory **132** (2012), no. 4, 543–564.

[Joc82] Naomi Jochnowitz, *Congruences between systems of eigenvalues of modular forms*, Trans. Amer. Math. Soc. **270** (1982), no. 1, 269–285.

[Joh17] Christian Johansson, *On the Sato-Tate conjecture for non-generic abelian surfaces*, Trans. Amer. Math. Soc. **369** (2017), no. 9, 6303–6325, With an appendix by Francesc Fité.

[Jon11] Owen T.R. Jones, *An analogue of the BGG resolution for locally analytic principal series*, Journal of Number Theory **131** (2011), no. 9, 1616 – 1640.

[JS81] Hervé Jacquet and Joseph A. Shalika, *On Euler products and the classification of automorphic forms. II*, Amer. J. Math. **103** (1981), no. 4, 777–815.

[Kas06] Payman L. Kassaei, *A gluing lemma and overconvergent modular forms*, Duke Math. J. **132** (2006), no. 3, 509–529.

[Kas16] ———, *Analytic continuation of overconvergent Hilbert modular forms*, Astérisque (2016), no. 382, 1–48.

[Kem78] George Kempf, *The Grothendieck–Cousin complex of an induced representation*, Advances in Mathematics **29** (1978), no. 3, 310 – 396.

[Kim03] Henry H. Kim, *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2, J. Amer. Math. Soc. **16** (2003), no. 1, 139–183, With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak.

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

[Kis09] ———, *Moduli of finite flat group schemes, and modularity*, Annals of Math.(2) **170** (2009), no. 3, 1085–1180.

[Kle94] Steven L. Kleiman, *The standard conjectures*, Motives (Seattle, WA, 1991), Proc. Sympos. Pure Math., vol. 55, Amer. Math. Soc., Providence, RI, 1994, pp. 3–20.

[Kot92] Robert E. Kottwitz, *Points on some Shimura varieties over finite fields*, J. Amer. Math. Soc. **5** (1992), no. 2, 373–444.

[KR99] Stephen S. Kudla and Michael Rapoport, *Arithmetic Hirzebruch-Zagier cycles*, J. Reine Angew. Math. **515** (1999), 155–244.

[KST14] Payman L. Kassaei, Shu Sasaki, and Yichao Tian, *Modularity lifting results in parallel weight one and applications to the Artin conjecture: the tamely ramified case*, Forum Math. Sigma **2** (2014), e18, 58.

[KT17] Chandrashekhar B. Khare and Jack A. Thorne, *Potential automorphy and the Leopoldt conjecture*, Amer. J. Math. **139** (2017), no. 5, 1205–1273.

[Lan80] Robert P. Langlands, *Base change for* GL(2), Annals of Mathematics Studies, vol. 96, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1980.

[Lan13] Kai-Wen Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, London Mathematical Society Monographs Series, vol. 36, Princeton University Press, Princeton, NJ, 2013.

[Lan16] ———, *Compactifications of PEL-type Shimura varieties in ramified characteristics*, Forum Math. Sigma **4** (2016), e1, 98.

[Lan17] ———, *Integral models of toroidal compactifications with projective cone decompositions*, International Mathematics Research Notices **11** (2017), 3237–3280.

[LPS13] Giuseppe Lombardo, Chris Peters, and Matthias Schütt, *Abelian fourfolds of Weil type and certain K3 double planes*, Rend. Semin. Mat. Univ. Politec. Torino **71** (2013), no. 3-4, 339–383.

[Maz89] Barry Mazur, *Deforming Galois representations*, Galois groups over **Q** (Berkeley, CA, 1987), Math. Sci. Res. Inst. Publ., vol. 16, Springer, New York, 1989, pp. 385–437.

[MB89] Laurent Moret-Bailly, *Groupes de Picard et problèmes de Skolem II*, Ann. Scient. Ec. Norm. Sup. **22** (1989), 181–194.

[MB90] ———, *Extensions de corps globaux à ramification et groupe de Galois donnés*, C. R. Acad. Sci. Paris Sér. I Math. **311** (1990), no. 6, 273–276.

[Mes72] William Messing, *The crystals associated to Barsotti-Tate groups: with applications to abelian schemes*, Lecture Notes in Mathematics, Vol. 264, Springer-Verlag, Berlin-New York, 1972.

[Mil79] James S. Milne, *Points on Shimura varieties mod p*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 165–184.

[Mok14] Chung Pang Mok, *Galois representations attached to automorphic forms on* GL_2 *over CM fields*, Compos. Math. **150** (2014), no. 4, 523–567.

[Mor84] David R. Morrison, *On K3 surfaces with large Picard number*, Invent. Math. **75** (1984), no. 1, 105–121.

[Mor85] ———, *The Kuga-Satake variety of an abelian surface*, J. Algebra **92** (1985), no. 2, 454–476.

[MT02] Abdellah Mokrane and Jacques Tilouine, *Cohomology of Siegel varieties with p-adic integral coefficients and applications*, Astérisque (2002), no. 280, 1–95, Cohomology of Siegel varieties.

[MT15] Chung Pang Mok and Fucheng Tan, *Overconvergent families of Siegel-Hilbert modular forms*, Canad. J. Math. **67** (2015), no. 4, 893–922.

[Mum08] David Mumford, *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, vol. 5, Published for the Tata Institute of Fundamental Research, Bombay; by Hindustan Book Agency, New Delhi, 2008, With appendices by C. P. Ramanujam and Yuri Manin, Corrected reprint of the second (1974) edition.

[MW16a] Colette Mœglin and Jean-Loup Waldspurger, *Stabilisation de la formule des traces tordue*, vol. 1, Springer International Publishing, 2016.

[MW16b] ———, *Stabilisation de la formule des traces tordue*, vol. 2, Springer International Publishing, 2016.

[MZ95] Ben J. J. Moonen and Yuri G. Zarhin, *Hodge classes and Tate classes on simple abelian fourfolds*, Duke Math. J. **77** (1995), no. 3, 553–581.

[Nak84] Shōichi Nakajima, *On Galois module structure of the cohomology groups of an algebraic variety*, Invent. Math. **75** (1984), no. 1, 1–8.

[NG02] B. C. Ngô and A. Genestier, *Alcôves et p-rang des variétés abéliennes*, Ann. Inst. Fourier (Grenoble) **52** (2002), no. 6, 1665–1680.

[Noo13] Rutger Noot, *The system of representations of the Weil-Deligne group associated to an abelian variety*, Algebra Number Theory **7** (2013), no. 2, 243–281.

[Noo17] ———, *The p-adic representation of the Weil-Deligne group associated to an abelian variety*, J. Number Theory **172** (2017), 301–320.

[Pat19] Stefan Patrikis, *Variations on a theorem of Tate*, Mem. Amer. Math. Soc. **258** (2019), no. 1238, viii+156.

[Pil11] Vincent Pilloni, *Prolongement analytique sur les variétés de Siegel*, Duke Math. J. **157** (2011), no. 1, 167–222.

[Pil12] ———, *Modularité, formes de Siegel et surfaces abéliennes*, J. Reine Angew. Math. **666** (2012), 35–82.

[Pil17] ———, *Formes modulaires p-adiques de Hilbert de poids 1*, Invent. Math. **208** (2017), no. 2, 633–676.

[Pil20] ———, *Higher coherent cohomology and p-adic modular forms of singular weights*, Duke Math. J. **169** (2020), no. 9, 1647–1807.

[Pin90] Richard Pink, *Arithmetical compactification of mixed Shimura varieties*, Bonner Mathematische Schriften [Bonn Mathematical Publications], vol. 209, Universität

Bonn, Mathematisches Institut, Bonn, 1990, Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, 1989.

[Poi08] Jérôme Poineau, *Un résultat de connexité pour les variétés analytiques p-adiques: privilège et noethérianité*, Compos. Math. **144** (2008), no. 1, 107–133.

[PS09] Ameya Pitale and Ralf Schmidt, *Bessel models for lowest weight representations of* $\mathrm{GSp}(4, \mathbf{R})$, Int. Math. Res. Not. IMRN (2009), no. 7, 1159–1212.

[PS16a] Vincent Pilloni and Benoît Stroh, *Cohomologie cohérente et représentations Galoisiennes*, Ann. Math. Qué. **40** (2016), no. 1, 167–202.

[PS16b] ———, *Surconvergence, ramification et modularité*, Astérisque (2016), no. 382, 195–266.

[PSY17] Cris Poor, Jerry Shurman, and David S. Yuen, *Siegel paramodular forms of weight 2 and squarefree level*, Int. J. Number Theory **13** (2017), no. 10, 2627–2652.

[PT15] Stefan Patrikis and Richard Taylor, *Automorphy and irreducibility of some l-adic representations*, Compos. Math. **151** (2015), no. 2, 207–229.

[PVZ16] Stefan Patrikis, José Felipe Voloch, and Yuri G. Zarhin, *Anabelian geometry and descent obstructions on moduli spaces*, Algebra Number Theory **10** (2016), no. 6, 1191–1219.

[PY15] Cris Poor and David S. Yuen, *Paramodular cusp forms*, Math. Comp. **84** (2015), no. 293, 1401–1438.

[Ram00] Dinakar Ramakrishnan, *Modularity of the Rankin-Selberg L-series, and multiplicity one for* $\mathrm{SL}(2)$, Ann. of Math. (2) **152** (2000), no. 1, 45–111.

[Ray94] Michel Raynaud, *1-motifs et monodromie géométrique*, Astérisque (1994), no. 223, 295–319, Périodes p-adiques (Bures-sur-Yvette, 1988).

[Rib04] Kenneth A. Ribet, *Abelian varieties over* $\mathbf{Q}$ *and modular forms*, Modular curves and abelian varieties, Progr. Math., vol. 224, Birkhäuser, Basel, 2004, pp. 241–261.

[Rie59] Bernhard Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie (1859), 671–680.

[Rob01] Brooks Roberts, *Global L-packets for* $\mathrm{GSp}(2)$ *and theta lifts*, Doc. Math. **6** (2001), 247–314 (electronic).

[RS07a] Dinakar Ramakrishnan and Freydoon Shahidi, *Siegel modular forms of genus 2 attached to elliptic curves*, Math. Res. Lett. **14** (2007), no. 2, 315–332.

[RS07b] Brooks Roberts and Ralf Schmidt, *Local newforms for GSp(4)*, Lecture Notes in Mathematics, vol. 1918, Springer, Berlin, 2007.

[RS07c] ———, *Tables for representations of GSp(4)*, available at `http://www2.math.ou.edu/~rschmidt/papers/tables.pdf`, 2007.

[RZ96] Michael Rapoport and Thomas Zink, *Period spaces for p-divisible groups*, Annals of Mathematics Studies, vol. 141, Princeton University Press, Princeton, NJ, 1996.

[Sai97] Takeshi Saito, *Modular forms and p-adic Hodge theory*, Invent. Math. **129** (1997), no. 3, 607–620.

[Sas13] Shu Sasaki, *On Artin representations and nearly ordinary Hecke algebras over totally real fields*, Doc. Math. **18** (2013), 997–1038.

[Saw16] William F. Sawin, *Ordinary primes for Abelian surfaces*, C. R. Math. Acad. Sci. Paris **354** (2016), no. 6, 566–568.

[SBT97] Nicholas I. Shepherd-Barron and Richard Taylor, mod 2 *and* mod 5 *icosahedral representations*, J. Amer. Math. Soc. **10** (1997), no. 2, 283–298.

[Sch15] Peter Scholze, *On torsion in the cohomology of locally symmetric varieties*, Ann. of Math. (2) **182** (2015), no. 3, 945–1066.

[Sch17] Ralf Schmidt, *Archimedean aspects of Siegel modular forms of degree 2*, Rocky Mountain J. Math. **47** (2017), no. 7, 2381–2422.

[Sch18] ———, *Packet structure and paramodular forms*, Trans. Amer. Math. Soc. **370** (2018), no. 5, 3085–3112.

[Sch19] Ciaran Schembri, *Examples of genuine QM abelian surfaces which are modular*, Res. Number Theory **5** (2019), no. 1, Paper No. 11, 12.

[Sch20] Ralf Schmidt, *Paramodular forms in CAP representations of* $\mathrm{GSp}(4)$, Acta Arith. **194** (2020), no. 4, 319–340.

[SD73] Henry Peter Francis Swinnerton-Dyer, *On l-adic representations and congruences for coefficients of modular forms*, Modular functions of one variable, III (Proc.

Internat. Summer School, Univ. Antwerp, 1972), Springer, Berlin, 1973, pp. 1–55. Lecture Notes in Math., Vol. 350.

[Ser68] Jean-Pierre Serre, *Abelian l-adic representations and elliptic curves*, McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam, 1968.

[Ser70] ———, *Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures)*, Séminaire Delange-Pisot-Poitou. Théorie des nombres **11** (1969-1970), no. 2, 1–15 (fre).

[Ser73] ———, *Formes modulaires et fonctions zêta p-adiques*, Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972), Springer, Berlin, 1973, pp. 191–268. Lecture Notes in Math., Vol. 350.

[Ser77] ———, *Modular forms of weight one and Galois representations*, Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975), Academic Press, London, 1977, pp. 193–268.

[Ser89] ———, *Lectures on the Mordell-Weil theorem*, Aspects of Mathematics, E15, Friedr. Vieweg & Sohn, Braunschweig, 1989, Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt.

[Ser00] ———, *Lettre à Marie-France Vignéras du 10/02/1987. Collected papers IV*, Springer-Verlag, Berlin, 2000, pp. 38–55.

[Sha97] Freydoon Shahidi, *On non-vanishing of twisted symmetric and exterior square L-functions for* $\mathrm{GL}(n)$, Pacific J. Math. (1997), no. Special Issue, 311–322, Olga Taussky-Todd: in memoriam.

[Sho18] Jack Shotton, *The Breuil-Mézard conjecture when $l \neq p$*, Duke Math. J. **167** (2018), no. 4, 603–678.

[Ski09] Christopher Skinner, *A note on the p-adic Galois representations attached to Hilbert modular forms*, Doc. Math. **14** (2009), 241–258.

[Sno09] Andrew Snowden, *On two dimensional weight two odd representations of totally real fields*, 2009, arXiv:0905.4266.

[Sor10] Claus M. Sorensen, *Galois representations attached to Hilbert-Siegel modular forms*, Doc. Math. **15** (2010), 623–670.

[ST93] Paul J. Sally, Jr. and Marko Tadić, *Induced representations and classifications for* $\mathrm{GSp}(2, F)$ *and* $\mathrm{Sp}(2, F)$, Mém. Soc. Math. France (N.S.) (1993), no. 52, 75–133.

[Sta13] The Stacks Project Authors, *Stacks Project*, `http://stacks.math.columbia.edu`, 2013.

[Ste04] William A. Stein, *Shafarevich–Tate groups of nonsquare order*, Modular curves and abelian varieties, Progr. Math., vol. 224, Birkhäuser, Basel, 2004, pp. 277–289.

[Str15] Benoît Stroh, *Mauvaise réduction au bord*, Astérisque (2015), no. 370, 269–304.

[Tat79] John Tate, *Number theoretic background*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 3–26.

[Tay02] Richard Taylor, *Remarks on a conjecture of Fontaine and Mazur*, J. Inst. Math. Jussieu **1** (2002), no. 1, 125–143.

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

[Tay08] ———, *Automorphy for some l-adic lifts of automorphic mod l Galois representations. II*, Pub. Math. IHES **108** (2008), 183–239.

[Tho12] Jack A. Thorne, *On the automorphy of l-adic Galois representations with small residual image*, J. Inst. Math. Jussieu **11** (2012), no. 4, 855–920, With an appendix by Robert Guralnick, Florian Herzig, Richard Taylor and Thorne.

[Tho15] ———, *Automorphy lifting for residually reducible l-adic Galois representations*, J. Amer. Math. Soc. **28** (2015), no. 3, 785–870.

[Til96] Jacques Tilouine, *Deformations of Galois representations and Hecke algebras*, Published for The Mehta Research Institute of Mathematics and Mathematical Physics, Allahabad, 1996.

[Til98] ———, *Deformations of Siegel-Hilbert Hecke eigensystems and their Galois representations*, Number theory (Tiruchirapalli, 1996), Contemp. Math., vol. 210, Amer. Math. Soc., Providence, RI, 1998, pp. 195–225.

[Til06a] ———, *Nearly ordinary rank four Galois representations and p-adic Siegel modular forms*, Compos. Math. **142** (2006), no. 5, 1122–1156, With an appendix by Don Blasius.

[Til06b] ———, *Siegel varieties and p-adic Siegel modular forms*, Doc. Math. (2006), no. Extra Vol., 781–817.

[Til09] ———, *Cohomologie des variétés de Siegel et représentations galoisiennes associées aux représentations cuspidales cohomologiques de* $\mathrm{GSp}_4(\mathbf{Q})$, Algèbre et théorie des nombres. Années 2007–2009, Publ. Math. Univ. Franche-Comté Besançon Algèbr. Theor. Nr., Lab. Math. Besançon, Besançon, 2009, pp. 99–114.

[Til12] J. Tilouine, *Formes compagnons et complexe BGG dual pour* GSp_4, Ann. Inst. Fourier (Grenoble) **62** (2012), no. 4, 1383–1436.

[TU95] Jacques Tilouine and Éric Urban, *Familles p-adiques à trois variables de formes de Siegel et de représentations galoisiennes*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 1, 5–10.

[TU99] ———, *Several-variable p-adic families of Siegel-Hilbert cusp eigensystems and their Galois representations*, Ann. Sci. École Norm. Sup. (4) **32** (1999), no. 4, 499–574.

[Tun81] Jerrold Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. (N.S.) **5** (1981), no. 2, 173–175.

[TW95] Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[TY07] Richard Taylor and Teruyoshi Yoshida, *Compatibility of local and global Langlands correspondences*, J. Amer. Math. Soc. **20** (2007), no. 2, 467–493 (electronic).

[Urb11] Éric Urban, *Eigenvarieties for reductive groups*, Ann. of Math. (2) **174** (2011), no. 3, 1685–1784.

[Vig98] Marie-France Vignéras, *Induced r-representations of p-adic reductive groups*, Selecta Mathematica **4** (1998), no. 4, 549–623.

[Vig05] ———, *Pro-p-Iwahori Hecke ring and supersingular* $\overline{\mathbf{F}}_p$*-representations*, Math. Ann. **331** (2005), no. 3, 523–556.

[VW13] Eva Viehmann and Torsten Wedhorn, *Ekedahl-Oort and Newton strata for Shimura varieties of PEL type*, Math. Ann. **356** (2013), no. 4, 1493–1550.

[Wal84] Nolan R. Wallach, *On the constant term of a square integrable automorphic form*, Operator algebras and group representations, Vol. II (Neptun, 1980), Monogr. Stud. Math., vol. 18, Pitman, Boston, MA, 1984, pp. 227–237.

[Wei52] André Weil, *Number-theory and algebraic geometry*, Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950, vol. 2, Amer. Math. Soc., Providence, R. I., 1952, pp. 90–100.

[Wie14] Gabor Wiese, *On Galois representations of weight one*, Doc. Math. **19** (2014), 689–707.

[Wil95] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Yos80] Hiroyuki Yoshida, *Siegel's modular forms and the arithmetic of quadratic forms*, Invent. Math. **60** (1980), no. 3, 193–248.

[Yos84] ———, *On Siegel modular forms obtained from theta series*, J. Reine Angew. Math. **352** (1984), 184–219.

[Yu08] Chia-Fu Yu, *Irreducibility and p-adic monodromies on the Siegel moduli spaces*, Adv. Math. **218** (2008), no. 4, 1253–1285.

[Yu11] ———, *Geometry of the Siegel modular threefold with paramodular level structure*, Proc. Amer. Math. Soc. **139** (2011), no. 9, 3181–3190.

Email address: gboxer@math.uchicago.edu

The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA

Email address: fcale@math.uchicago.edu

The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA

Email address: toby.gee@imperial.ac.uk

Department of Mathematics, Imperial College London, London SW7 2AZ, UK

Email address: vincent.pilloni@ens-lyon.fr

Unité de Mathématiques pures et appliquées, Ecole normale supérieure de Lyon, 46 allée d'Italie, 69 364 Lyon Cedex 07, France