Journal of the Inst. of Math. Jussieu (2008) **7**(2), 247–268 © Cambridge University Press 247
doi:10.1017/S1474748007000199 Printed in the United Kingdom

METHODS FOR p-ADIC MONODROMY

CHING-LI CHAI

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19003, USA (chai@math.upenn.edu)

(Received 21 July 2006; accepted 20 January 2007)

Abstract We explain three methods for showing that the p-adic monodromy of a modular family of abelian varieties is 'as large as possible', and illustrate them in the case of the ordinary locus of the moduli space of g-dimensional principally polarized abelian varieties over a field of characteristic p. The first method originated from Ribet's proof of the irreducibility of the Igusa tower for Hilbert modular varieties. The second and third methods both exploit Hecke correspondences near a hypersymmetric point, but in slightly different ways. The third method was inspired by work of Hida, plus a group theoretic argument for the maximality of ℓ-adic monodromy with $\ell \neq p$.

Keywords: Shimura variety; Hecke correspondence; modular variety; monodromy; abelian variety

AMS 2000 *Mathematics subject classification:* Primary 14G35; 14G32; 11G18; 11F60

1. Introduction

Let p be a prime number, fixed throughout this note. Our central question is how to show that the p-adic monodromy of a modular family of abelian varieties is large.

Some clarification is called for.

- A 'modular family of abelian varieties' will be interpreted as a subvariety Z of a modular variety $\mathcal{M}$ of PEL type over an algebraically closed field $k \supset \mathbb{F}_p$ defined by fixing some invariant for geometric fibres of the Barsotti–Tate group $A[p^\infty] \to \mathcal{M}$ with prescribed symmetries attached to the universal abelian scheme $A \to \mathcal{M}$.

- Example of such invariants include the p-rank, Newton polygon, or the isomorphism type of the Barsotti–Tate group with prescribed symmetries. Typically such a subvariety Z is stable under all prime-to-p Hecke correspondences on $\mathcal{M}$; moreover any two geometric fibres of $A[p^\infty] \to Z$ are isogenous via a quasi-isogeny which preserves the prescribed endomorphisms and polarizations.

- The étale sheaf of such quasi-isogenies gives rise to a homomorphism $\rho_p = \rho_{p,Z}$ from the fundamental group of Z to the group of $\mathbb{Q}_p$-points of a linear algebraic group G over $\mathbb{Q}_p$, defined up to conjugation. Often the target of the p-adic monodromy homomorphism ρ_p is an open subgroup of $G(\mathbb{Q}_p)$. We abuse notation and denote this target group by $G(\mathbb{Z}_p)$.

- The expectation is that the p-adic monodromy is 'large', or 'maximal'. In other words, the image of ρ is expected to be equal to the target group $G(\mathbb{Z}_p)$, or equal to an open subgroup of $G(\mathbb{Z}_p)$ if we are less ambitious.

The first example of maximality of p-adic monodromy is a theorem of Igusa in [**18**], when Z is the open dense subset of the modular curve corresponding to ordinary elliptic curves. See part (i) of Remark 2.3 for the statement of Igusa's theorem and part (ii) for comments on generalizations of Igusa's method based on local p-adic monodromy. For generalizations to higher-dimensional modular varieties, the best-known examples in chronological order are the ordinary locus of a Hilbert modular variety and the ordinary locus of a Siegel modular variety.

In the second example, Z is the open dense subset of a Hilbert modular variety $\mathcal{M}_F$ attached to a totally real number field F, which classifies ordinary abelian varieties with endomorphisms by $\mathcal{O}_F$. The target of the p-adic monodromy homomorphism $\rho_{p,F}$ is the group $(\mathcal{O}_F \otimes_{\mathbb{Z}} \mathbb{Z}_p)^\times$ of local units. Ribet showed that $\rho_{p,F}$ is surjective (see [**24**] and [**10**]). Ribet's method in [**24**] and [**10**] is global and arithmetic in nature; it uses Frobenii attached to points over finite fields of the moduli space $\mathcal{M}_F$.

In the third example, Z is the ordinary locus of a Siegel modular variety $\mathcal{A}_{g,n}$, and the p-adic monodromy is equal to $\mathrm{GL}_g(\mathbb{Z}_p)$. See Theorem 2.1 for the precise statement and [**11**] and [**12**] for proofs. The proofs of Theorem 2.1 in [**11**] and [**12**] are based on considerations of local p adic monodromy (see Remark 2.3).

In this article we explain three methods for proving the maximality of p-adic monodromy. Instead of pushing for the most general case with each method, we choose to illustrate the methods for the third example above, when Z is the ordinary locus in the Siegel modular variety $\mathcal{A}_{g,n}$. In other word, we offer three proofs* of Theorem 2.1. Each method can be applied to more general situations, such as a leaf or a Newton polygon stratum in a modular variety of PEL-type. See § 6 for the case of a leaf in a Siegel modular or the ordinary locus of a modular variety of quasi-split PEL-type $\mathrm{U}(n, n)$. See also [**7**, § 5] for more information about the irreducibility of non-supersingular leaves and the maximality of their p-adic monodromy in the case of Siegel modular varieties.

The first of the three proofs of Theorem 2.1 generalizes of an argument of Ribet in [**24**] and [**10**] to the situation when the target of the p-monodromy representation is noncommutative. The other two proofs follow a common thread of ideas, in that they both use Hecke correspondences with a given hypersymmetric point (in the sense of [**6**]) as a fixed point; these Hecke correspondences form the *local stabilizer subgroup* H_{x_0} of the given hypersymmetric point. In the second proof one applies the local stabilizer subgroup H_{x_0} of a hypersymmetric point x_0 to a modular subvariety $B \ni x_0$ with known p-adic monodromy to produce many subvarieties of Z with known p-adic monodromy. In the third proof one examines the action of the local stabilizer subgroup of a hypersymmetric point x_0 on a tower of finite étale covers of Z which defines the p-adic monodromy representation ρ_Z; the result is that the image of the local stabilizer subgroup H_{x_0} is contained in the image of the p-adic monodromy. The second proof was sketched in

* In the second proof we assume that $p > 2$.

[**5**]. The third proof was inspired by Hida's work on p-adic monodromy in [**14**] (see also [**16**]) and drew on an argument in [**4**]. The three proofs are explained in §§ 3, 4 and 5, respectively. Sketches of the ideas of the proofs can be found at the beginning of these sections. In § 6 we indicate how the methods in this article can be applied to show the maximality of p-adic monodromy groups in various situations, including the ordinary locus of the modular variety of quasi-split PEL-type $\mathrm{U}(n,n)$ and leaves in Siegel modular varieties.

If we compare the above three methods for proving maximality of p-adic monodromy, the second and the third methods using Hecke correspondences have the aura of a 'pure-thought proof'; this is especially true for the third method. Since the last two methods depend on the existence of 'hypersymmetric points' in Z, there are situations when they do not apply, while Ribet's method is not burdened by such restrictions (see Remark 6.1). Together these three methods support the contention that '*p-adic monodromy of a modular family of abelian varieties is easy*'.*

2. Notation

2.1. Let p be a prime number. Let $\mathbb{A}_f^{(p)} = \prod'_{\ell\neq p} \mathbb{Q}_\ell$ be the ring of finite prime-to-p adeles. Denote by $\mathbb{Z}_{(p)}$ the subring of $\mathbb{Q}$ consisting of all rational numbers whose denominator is prime to p.

Let $n \geqslant 3$ be a positive integer, $(n,p) = 1$. Let $\mathcal{A}_{g,n}^{\mathrm{ord}}$ be the moduli space over $\overline{\mathbb{F}}_p$ of g-dimensional ordinary principally polarized abelian varieties in characteristic p with symplectic level-n structure.

Let $A[p^\infty]_{\mathrm{et}} \to \mathcal{A}_{g,n}^{\mathrm{ord}}$ be the maximal étale quotient of the Barsotti–Tate group $A[p^\infty] \to \mathcal{A}_{g,n}^{\mathrm{ord}}$ attached to the universal abelian scheme $A \to \mathcal{A}_{g,n}^{\mathrm{ord}}$; it is an étale Barsotti–Tate group of height g over $\mathcal{A}_{g,n}^{\mathrm{ord}}$.

Let $x_0 = [(A_0, \lambda_0, \eta_0)]$ be an $\overline{\mathbb{F}}_p$-point of $\mathcal{A}_{g,n}^{\mathrm{ord}}$, where λ_0 is a principal polarization of an ordinary abelian variety A_0 over $\overline{\mathbb{F}}_p$, and η_0 is a level-n structure on A_0. Let $\mathrm{T}_0 = \mathrm{T}_p(A_0[p^\infty]_{\mathrm{et}})$ be the p-adic Tate module attached to the maximal étale quotient $A_0[p^\infty]_{\mathrm{et}}$ of the Barsotti–Tate group $A_0[p^\infty]$ attached to A_0. Notice that $\mathrm{T}_0 \cong \mathbb{Z}_p^g$ non-canonically.

Let

$$\rho_p = \rho_{p,\mathcal{A}_{g,n}^{\mathrm{ord}}} : \pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0) \to \mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_g(\mathbb{Z}_p)$$

be the p-adic monodromy representation defined by the base point x_0 of $\mathcal{A}_{g,n}^{\mathrm{ord}}$. If we view $\pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0)$ as a quotient of the Galois group $G_{\mathcal{A}_{g,n}}$ of the function field of $\mathcal{A}_{g,n}^{\mathrm{ord}}$, then ρ_p corresponds to the limit of the natural actions of $G_{\mathcal{A}_{g,n}}$ on the generic fibre of the finite étale group schemes $A[p^m]_{\mathrm{et}} \to \mathcal{A}_{g,n}^{\mathrm{ord}}$, $m \in \mathbb{N}$. See § 2.4 for another definition of p-adic monodromy.

Theorem 2.1. *The image of the p-adic monodromy homomorphism* $\rho_{p,\mathcal{A}_{g,n}^{\mathrm{ord}}}$ *is equal to* $\mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_g(\mathbb{Z}_p)$.

* In contrast, the important question on the semisimplicity of p-adic monodromy of an 'arbitrary' family of abelian varieties in characteristic p seems inaccessible at present.

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

Remark 2.2. The special case of Theorem 2.1 when $g = 1$ is a classical theorem of Igusa in [**18**]. See Theorem 4.3 on p. 149 of [**19**] for an exposition of Igusa's theorem.

Remark 2.3.

(i) Igusa's proof uses the local monodromy at a point of $\mathcal{A}_{1,n}$ which corresponds to a supersingular elliptic curve. In this case the image of the local p-adic monodromy is already equal to the target group $\mathbb{Z}_p^\times$ of the p-adic monodromy representation.

(ii) Ekedahl's proof in [**11**] shows that the p-adic local monodromy at a superspecial point of $\mathcal{A}_{g,n}$ (i.e. a point which corresponds to the product of g copies of a supersingular elliptic curve) is already equal to $\mathrm{GL}_g(\mathbb{Z}_p)$. This is an exact generalization of Igusa's proof. In [**11**] Ekedahl used curves of genus two instead of abelian varieties, but it is clear that one can also use deformation theory of abelian varieties. See [**21**] for the case of Picard modular varieties, and [**1**] for deformations of p-divisible groups with large local p-adic monodromy. There is one disadvantage of the Igusa–Ekedahl method: substantial effort is required when one uses this method to compute the p-adic monodromy of a subvariety Z defined by p-adic properties of the universal Barsotti–Tate groups. The work is in constructing explicit local coordinates of this subvariety Z at a *basic point* z and the computation of the Galois group of suitable finite extensions of the function field of the formal completion of Z at z.

(iii) The proof in [**12**] is also based on local monodromy. It uses the arithmetic compactification theory to show that the p-adic local monodromy at a zero-dimensional cusp of the minimal compactification of $\mathcal{A}_{g,n}$ is equal to $\mathrm{SL}_g(\mathbb{Z}_p)$. This method applies to fewer situations for two reasons. First, the modular variety $\mathcal{M}$ may be proper, i.e. the boundary of $\mathcal{M}$ may be empty. Even when the modular variety has a boundary, the Zariski closure of the modularly defined subvariety Z may not intersect the boundary of $\mathcal{M}$. Secondly, when the boundary of $\mathcal{M}$ is not empty, the local monodromy at the boundary may be still be 'too small'. For instance in the case of a Hilbert modular variety $\mathcal{M}_{F,n}$ over $\overline{\mathbb{F}}_p$ with $(n, p) = 1$, the local monodromy at a cusp in the minimal compactification of $\mathcal{M}_{F,n}$ is the p-adic completion of the subgroup of units in $\mathcal{O}_F^\times$ which are congruent to 1 modulo n.* Notice that the image of the local monodromy at a cusp is contained in the subgroup $\mathrm{Ker}(\mathrm{N}_{F/\mathbb{Q}} : (F \otimes \mathbb{Q}_p)^\times \to \mathbb{Q}_p^\times)$ of the target group $(F \otimes \mathbb{Q}_p)^\times$ of ρ_p; this $([F : \mathbb{Q}]-1)$-dimensional subgroup is the target group of the local p-adic monodromy. *A priori*, the dimension of the local p-adic monodromy group at the cusps of a Hilbert modular variety $\mathcal{M}_{F,n}$ may be smaller than $[F : \mathbb{Q}] - 1$, the dimension of its target; it is equal to $[F : \mathbb{Q}] - 1$ if and only if the Leopoldt conjecture holds for F!

2.2. In the rest of this article, we will take the base point $x_0 = [(A_0, \lambda_0, \eta_0)]$ of $\mathcal{A}_{g,n}^{\mathrm{ord}}$ to be of the form $A_0 = E_1 \times_{\mathrm{Spec}(\overline{\mathbb{F}}_p)} \cdots \times_{\mathrm{Spec}(\overline{\mathbb{F}}_p)} E_1$, where E_1 is an ordinary elliptic curve

* This is what the argument in [**2**] shows. The statement in [**2**] about the p-adic monodromy of a Hilbert modular variety is wrong.

over $\overline{\mathbb{F}}_p$, and λ_0 is the g-fold product of the principal polarization of E_1. Let $E_1[p^\infty]_{\mathrm{mult}}$ (respectively, $E_1[p^\infty]_{\mathrm{et}}$) be the multiplicative part of $E_1[p^\infty]$ (respectively, the maximal étale quotient of $E_1[p^\infty]$). Let $\mathrm{T} := \mathrm{T}_p(E_1[p^\infty]_{\mathrm{et}})$, which is a free $\mathbb{Z}_p$-module of rank one. Then $\mathrm{T}_0 := \mathrm{T}_p(A_0[p^\infty]_{\mathrm{et}})$ is the direct sum of g copies of T, and $\mathrm{GL}_{\mathbb{Z}_p}(\mathrm{T}_0)$ is canonically isomorphic to $\mathrm{GL}_g(\mathrm{End}(\mathrm{T})) = \mathrm{GL}_g(\mathbb{Z}_p)$.

Let $\mathcal{O} = \mathrm{End}(E_1)$ and $K = \mathrm{End}(E_1) \otimes_{\mathbb{Z}} \mathbb{Q}$. So $\mathrm{End}(A_0)$ (respectively, $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q}$) is canonically isomorphic to $\mathrm{M}_g(\mathcal{O})$ (respectively, $\mathrm{M}_g(K)$). It is well known that K is an imaginary quadratic field, and $\mathcal{O}$ is an order of K such that $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p \cong \mathbb{Z}_p \times \mathbb{Z}_p$. The two factors of $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ correspond to the action of $\mathcal{O}$ on $E_1[p^\infty]_{\mathrm{mult}}$ and $E_1[p^\infty]_{\mathrm{et}}$ respectively. Via the above isomorphisms, the Rosati involution $*$ attached to the principal polarization of λ_0 corresponds to the involution $C \mapsto {}^{\mathrm{t}}\bar{C}$ on $\mathrm{M}_g(\mathcal{O})$ and $\mathrm{M}_g(K)$. On $\mathrm{M}_g(\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p) = \mathrm{M}_g(\mathbb{Z}_p) \times \mathrm{M}_g(\mathbb{Z}_p)$, the Rosati involution is $(C_1, C_2) \mapsto ({}^{\mathrm{t}}C_2, {}^{\mathrm{t}}C_1)$, $(C_1, C_2) \in \mathrm{M}_g(\mathbb{Z}_p) \times \mathrm{M}_g(\mathbb{Z}_p)$.

Denote by H the unitary group attached to the semisimple $\mathbb{Q}$-algebra $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathrm{M}_g(K)$ and the involution $*$. The reductive linear algebraic group H over $\mathbb{Q}$ is characterized by the property that

$$\mathrm{H}(R) = \{x \in (\mathrm{End}(A_0) \otimes_{\mathbb{Z}} R)^\times \mid x \cdot *(x) = *(x) \cdot x = 1\}.$$

Under the natural isomorphism $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathrm{M}_g(K)$ and the isomorphism

$$\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q}_p \cong \mathrm{M}_g(\mathbb{Q}_p) \times \mathrm{M}_g(\mathbb{Q}_p)$$

induced by $K \otimes_{\mathbb{Q}} \mathbb{Q}_p \xrightarrow{\sim} \mathbb{Q}_p \times \mathbb{Q}_p$, $\mathrm{H}(\mathbb{Q})$ is identified with the set of all matrices $C \in \mathrm{M}_g(K)$ such that $C \cdot *(C) = *(C) \cdot C = \mathrm{Id}_g$, and $\mathrm{H}(\mathbb{Q}_p)$ is identified with the subset of all pairs $(C_1, C_2) \in \mathrm{M}_g(\mathbb{Q}_p) \times \mathrm{M}_g(\mathbb{Q}_p)$ such that $C_1 \cdot {}^{\mathrm{t}}C_2 = \mathrm{Id}_g = C_2 \cdot {}^{\mathrm{t}}C_1$. The second projection from $\mathrm{M}_g(\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p)$ to $\mathrm{M}_g(\mathbb{Z}_p)$ induces an isomorphism $\mathrm{pr} : \mathrm{H}(\mathbb{Q}_p) \xrightarrow{\sim} \mathrm{GL}_g(\mathbb{Q}_p)$. We abuse notation and write $\mathrm{H}(\mathbb{Z}_p)$ for the compact open subgroup $\mathrm{pr}^{-1}(\mathrm{GL}_g(\mathbb{Z}_p))$; it is the set of all elements $\beta \in \mathrm{H}(\mathbb{Q}_p)$ such that β induces an automorphism of $A_0[p^\infty]$. Denote by $\mathrm{H}(\mathbb{Z}_{(p)})$ the subgroup $\mathrm{H}(\mathbb{Q}) \cap \mathrm{H}(\mathbb{Z}_p)$ of $\mathrm{H}(\mathbb{Q})$. The image of $\mathrm{H}(\mathbb{Z}_{(p)})$ in $\mathrm{H}(\mathbb{Z}_p)$ is a dense subgroup of $\mathrm{H}(\mathbb{Z}_p)$ for the p-adic topology.

2.3. Two features of the base point x_0 deserve attention.

(1) The abelian variety A_0 is *hypersymmetric* in the sense that

$$\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q}_p \xrightarrow{\sim} \mathrm{End}(A_0[p^\infty]) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p.$$

See [**6**] for more discussion on the notion of hypersymmetric abelian varieties.

(2) The order $\mathrm{End}(A_0)$ of the semisimple $\mathbb{Q}$-algebra $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q}$ is maximal at p. In view of (1) above, this means that $\mathrm{End}(A_0[p^\infty])$ is a maximal order of the semisimple $\mathbb{Q}_p$-algebra $\mathrm{End}(A_0[p^\infty]) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.

2.4. Remarks on the definition of p-adic monodromy

Let $(B,\mu)\to S$ be a principally polarized abelian scheme over an irreducible normal $\mathbb{F}_p$-scheme S such that all fibres of B are ordinary abelian varieties. Let s_0 be a geometric point of S. When B is ordinary, one can define the p-adic monodromy of $B\to S$ to be the homomorphism $\pi_1(S,s_0)\to \mathrm{GL}_{\mathbb{Z}_p}(\mathrm{T}_p(B_{s_0}[p^\infty]_{\mathrm{et}}))$ attached to the smooth $\mathbb{Z}_p$-sheaf $\mathrm{T}_p(B[p^\infty]_{\mathrm{et}})$ over S, where B_{s_0} is the fibre of $B\to S$ at s_0. There is an equivalent definition which works for more general situations, such as when $(B,\mu)[p^\infty]\to S$ is fibrewise geometrically constant. Let $G:=\mathrm{Aut}((B_{s_0},\mu_{s_0})[p^\infty])$. Consider the sheaf

$$\mathcal{I}:=\underline{\mathrm{Isom}}_{S^{\mathrm{perf}}}((B_{s_0},\mu_{s_0})[p^\infty],(B,\mu)[p^\infty]),$$

which is a right G-torsor over the perfection S^{perf} of S. Notice that $B[p^\infty]$ splits uniquely over S^{perf} as the direct product of its maximal toric part $B[p^\infty_{\mathrm{toric}}]$ and its maximal étale quotient $B[p^\infty_{\mathrm{et}}]$. Then the p-adic monodromy $\rho_{B/S}:\pi_1(S,s_0)\to G$ is the homomorphism from $\pi_1(S^{\mathrm{perf}},s_0)=\pi_1(S,s_0)$ to the profinite p-adic Lie group G attached to the G-torsor $\mathcal{I}$.

We indicate why the two definitions are equivalent. Let $\mathcal{I}'$ be the subsheaf of the étale sheaf $\underline{\mathrm{Isom}}_S(B_{s_0}[p^\infty]_{\mathrm{toric}},B[p^\infty]_{\mathrm{toric}})\times_S\underline{\mathrm{Isom}}_S(B_{s_0}[p^\infty]_{\mathrm{et}},B[p^\infty]_{\mathrm{et}})$ which are compatible with the polarizations $\mu_{s_0}[p^\infty]$ and $\mu[p^\infty]$. Let $\mathcal{I}'_{\mathrm{et}}:=\underline{\mathrm{Isom}}_S(B_{s_0}[p^\infty]_{\mathrm{et}},B[p^\infty]_{\mathrm{et}})$. The subgroup G' consisting of all elements of $\mathrm{Aut}(B_{s_0}[p^\infty]_{\mathrm{toric}})\times\mathrm{Aut}(B_{s_0}[p^\infty]_{\mathrm{et}})$ which are compatible with the polarizations is naturally identified with G.

The projection $G'\to\mathrm{GL}(\mathrm{T}_p(B_{s_0}[p^\infty]_{\mathrm{et}}))$ is an isomorphism because the principal polarization μ_{s_0} identifies $B_{s_0}[p^\infty]_{\mathrm{toric}}$ as the dual of $B_{s_0}[p^\infty]_{\mathrm{et}}$. Similarly, the natural projection map $\mathcal{I}'\to\mathcal{I}'_{\mathrm{et}}$ is an isomorphism. Moreover, the homomorphisms $\rho_{\mathcal{I}'}:\pi_1(S,s_0)\to G'\cong G$ and $\rho_{\mathcal{I}'_{\mathrm{et}}}:\pi_1(S,s_0)\to\mathrm{Aut}(B_{s_0}[p^\infty]_{\mathrm{et}})\cong G$ attached to $\mathcal{I}'$ and $\mathcal{I}'_{\mathrm{et}}$ are equal. Notice that the homomorphism $\rho_{\mathcal{I}'}:\pi_1(S,s_0)\to G'\xrightarrow{\sim}\mathrm{GL}(\mathrm{T}_p(B_{s_0}[p^\infty]_{\mathrm{et}}))$ is nothing but the p-adic monodromy representation coming from the action of the Galois group of the function field $\kappa(S)$ of S on the p-power torsion points $B[p^\infty](\kappa(S)^{\mathrm{alg}})$. On the other hand, because the étale topology is insensitive to nilpotent extensions, the homomorphisms $\rho_{\mathcal{I}}:\pi_1(S,s_0)\to G$ and $\rho_{\mathcal{I}'}:\pi_1(S,s_0)\to G$ are equal. So the two definitions are equivalent.

3. Ribet's method revisited

3.1. Sketch of idea

The goal of this section is to prove Theorem 2.1 using Ribet's method in [**10**, **24**]. Since the target of the p-adic monodromy homomorphism for the ordinary locus of Hilbert modular varieties is commutative, it was often thought that Ribet's method would have difficulty producing information beyond the abelianized p-adic monodromy. This is not the case at all. Indeed Ribet's method can be used to compute the p-adic monodromy of leaves in a Hilbert modular variety $\mathcal{M}_{F,n}$ for instance, where the target of the p-adic monodromy homomorphism may be non-commutative (see [**7**, Theorem 4.5]).

The proof of Theorem 2.1 in this section consists of a few lemmas in group theory, followed by the body of the proof in § 3.3. For instance Lemma 3.2 says that a suitable congruence condition modulo p^2 on the characteristic polynomial of a semisimple conjugacy class in $\mathrm{GL}_2(\mathbb{Z}_p)$ ensures that the reduction modulo p of such a semisimple conjugacy class is a non-trivial unipotent conjugacy class in $\mathrm{GL}_2(\mathbb{F}_p)$. According to these lemmas, we only need to show that the image of ρ contains certain congruent classes modulo p^2, in the case $g = 2$.

The basic idea of the proof is as follows. Suppose we have two g-dimensional ordinary principally polarized abelian varieties A_1, A_2 over a common finite field $\mathbb{F}_q$, then the 'difference' $\mathrm{Fr}_{A_1} \cdot \mathrm{Fr}_{A_2}^{-1}$ gives an element of the geometric fundamental group $\pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0)$, and its image under the p-adic monodromy homomorphism ρ_p gives a conjugacy class of $\mathrm{GL}_g(\mathbb{Z}_p)$. See Remark 3.5 for further discussion about Frobenius at closed points. Of course there is ambiguity when one tries to form the difference between two conjugacy classes, but if we restrict ourselves to the case when the action of Fr_{A_2} modulo p^N lies in the centre of $\mathrm{GL}_g(\mathbb{Z}/p^N\mathbb{Z})$, then the difference gives a well-defined conjugacy class in $\mathrm{GL}_g(\mathbb{Z}_p)$ modulo p^N.

There is a technical difference with [**24**] and [**10**]: we use Honda–Tate [**17**, **26**–**28**] to produce abelian varieties over finite fields with real multiplications, instead of using [**8**], which depends on Honda–Tate. One reason for this choice is to emphasize that Ribet's method applies to more general situations where the abelian varieties involved may not be ordinary.

3.2. Reduction to the case $g = 2$

The notation here is as in § 2.1. Recall that the target of the p-adic monodromy ρ is $\mathrm{GL}(\mathrm{T}_p(A_0[p^\infty])) = \mathrm{GL}_g(\mathrm{End}(\mathrm{T})) = \mathrm{GL}_g(\mathbb{Z}_p)$. The group $\mathrm{GL}_g(\mathbb{Z}_p)$ contains many copies of 'standardly embedded' $\mathrm{GL}_2(\mathbb{Z}_p)$ in block form. For any $1 \leqslant i_0 < j_0 \leqslant g$, the associated standardly embedded $\mathrm{GL}_2(\mathbb{Z}_p)$ consists of all elements $(a_{ij}) \in \mathrm{GL}_g(\mathbb{Z}_p)$ such $a_{ii} = 1$ if $i \neq i_0, j_0$, and $a_{ij} = 0$ if $i \neq j$ and $(i,j) \neq (i_0,j_0),(j_0,i_0)$; denote by H_J this standardly embedded subgroup of $\mathrm{GL}_g(\mathbb{Z}_p)$. In Lie theory these 'standardly embedded' SL_2 are called 'root subgroups'. It is easy to see that the $g-1$ standardly embedded subgroups $H_{\{1,2\}}, \dots, H_{\{g-1,g\}}$ of $\mathrm{GL}_g(\mathbb{Z}_p)$ generate $\mathrm{GL}_g(\mathbb{Z}_p)$ for any $g \geqslant 2$. A standardly embedded $\mathrm{GL}_2(\mathbb{Z}_p)$ inside $\mathrm{GL}_g(\mathbb{Z}_p)$ can be realized in terms of geometry of moduli spaces as follows. If we fix a level-n structure for the elliptic curve E_1 over $\overline{\mathbb{F}}_p$, then for each subset $J \subset \{1, 2, \dots, g\}$ with two elements of the form $J = \{j, j+1\}$ where j is an integer with $1 \leqslant j \leqslant g-1$, we have an embedding $i_J : \mathcal{A}_{2,n} \hookrightarrow \mathcal{A}_{g,n}$ such that the image of $\pi_1(\mathcal{A}_{2,n}^{\mathrm{ord}}, s_2)$ under $\rho_p : \pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0) \to \mathrm{GL}_g(\mathbb{Z}_p)$ is contained in the subgroup of J-blocks of $\mathrm{GL}_g(\mathbb{Z}_p)$ which is isomorphic to $\mathrm{GL}_2(\mathbb{Z}_p)$. Here s_2 denotes the $\overline{\mathbb{F}}_p$-point of $\mathcal{A}_{2,n}^{\mathrm{ord}}$ which corresponds to the abelian surface $E_1 \times_{\mathrm{Spec}(\overline{\mathbb{F}}_p)} E_1$ with the product principal polarization $\mu_1 \times \mu_1$ and product level-n structure $\eta_1 \times \eta_1$. This embedding i_J is defined by the family $(E_1, \mu_1, \eta_1)^{j-1} \times (A, \lambda, \eta) \times (E_1, \mu_1, \eta_1)^{g-j-1}$ over $\mathcal{A}_{2,n}$, where $(A, \lambda, \eta) \to \mathcal{A}_{2,n}$ denotes the universal principally polarized abelian scheme with level-n structure over $\mathcal{A}_{2,n}$. It is clear that the image of the composition

$$\pi_1(\mathcal{A}_{2,n}^{\mathrm{ord}}, s_2) \to \pi_1(\mathcal{A}_{g,n}, x_0) \xrightarrow{\rho} \mathrm{GL}_g(\mathbb{Z}_p)$$

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

is contained in the J-block subgroup $\mathrm{GL}_g(\mathbb{Z}_p)$, and this image subgroup is naturally isomorphic to the image of the p-adic monodromy representation $\pi_1(\mathcal{A}^{\mathrm{ord}}_{2,n}, s_2) \to \mathrm{GL}_2(\mathbb{Z}_p)$. Therefore, it suffices to prove Theorem 2.1 in the case when $g = 2$. This reduction step to the case $g = 2$ using the 'standardly embedded copies of $\mathcal{A}^{\mathrm{ord}}_{2,n}$ in $\mathcal{A}^{\mathrm{ord}}_{g,n}$' already appeared in [**11**].

Remark. Although not needed in this article, we note that it is possible to define the 'standard embedding' $i_J : \mathcal{A}_{2,n} \hookrightarrow \mathcal{A}_{g,n}$ for *any* subset $J \subseteq \{1, \dots, g\}$ with two elements, as follows. Suppose that $J = \{i_0, j_0\}$, $1 \leqslant i_0 < j_0 \leqslant g$. Consider the principally polarized abelian scheme $(B, \lambda_B, \eta_B) := (A, \lambda, \eta) \times (E_1, \mu_1, \eta_1)^{g-2}$ over $\mathcal{A}_{2,n}$. Change the level-n structure $\eta_B : B[n] \xrightarrow{\sim} (\mathbb{Z}/n\mathbb{Z})^{2n} = (\mathbb{Z}/n\mathbb{Z})^n \times (\mathbb{Z}/n\mathbb{Z})^n$ to $\tau_{\{1,2\},\{i_0,j_0\}}$, where $\tau_{\{1,2\},\{i,j\}}$ is the symplectic automorphism of $(\mathbb{Z}/n\mathbb{Z})^n \times (\mathbb{Z}/n\mathbb{Z})^n$ induced by the permutation $\sigma_{\{1,2\},\{i_0,j_0\}}$ of $\{1, \dots, n\}$ which sends 1 to i_0, 2 to j_0, and $\sigma_{\{1,2\}}(i) < \sigma_{\{1,2\}}(j)$ if $3 \leqslant i < j \leqslant g$. The resulting triple $(B, \lambda_B, \tau_{\{1,2\},\{i_0,j_0\}} \circ \eta_B)$ defines the standard embedding i_J.

Lemma 3.1. *Denote by Δ the subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$, isomorphic to $\mathbb{Z}_p^\times \times \mathbb{Z}_p^\times$, consisting of all diagonal matrices in $\mathrm{GL}_2(\mathbb{Z}_p)$. Let $\pi_0 : \mathrm{GL}_2(\mathbb{Z}_p) \to \mathrm{GL}_2(\mathbb{F}_p)$ be the natural surjection given by reduction modulo p. Let H be a closed subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$ which contains the subgroup Δ such that $\pi_0(H) = \mathrm{GL}_2(\mathbb{F}_p)$. Then $H = \mathrm{GL}_2(\mathbb{Z}_p)$.*

Proof. Denote by D the subgroup of $\mathrm{M}_2(\mathbb{F}_p)$ consisting of all diagonal 2×2 matrices with entries in $\mathbb{F}_p$. Denote by U_m the subgroup $1 + p^m\mathrm{M}_2(\mathbb{Z}_p)$ of $\mathrm{GL}_2(\mathbb{Z}_p)$ for every positive integer m. A simple calculation show that $\mathrm{M}_2(\mathbb{F}_p)$ is generated by $\mathrm{Ad}(\mathrm{GL}_2(\mathbb{F}_p)) \cdot D$, the set of all $\mathrm{GL}_2(\mathbb{F}_p)$-conjugates of D. The last statement implies that the composition

$$H \cap (1 + p^m\mathrm{M}_2(\mathbb{Z}_p)) \hookrightarrow U_m \to U_m/U_{m+1}$$

is surjective for every $m > 0$. Hence H surjects to $\mathrm{GL}_2(\mathbb{Z}/p^m\mathbb{Z})$ for every $m > 0$, and Lemma 3.1 follows. □

Lemma 3.2. *Let A be an element of $\mathrm{GL}_2(\mathbb{Z}_p)$ such that $\mathrm{tr}(A) \equiv 2 \pmod{p^2}$, $\det(A) \equiv 1 \pmod{p}$ and $\det(A) \not\equiv 1 \pmod{p^2}$. Then the image $\bar{A}$ of A in $\mathrm{GL}_2(\mathbb{F}_p)$ is conjugate to $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. In other words, $\bar{A}$ is a non-trivial unipotent element in $\mathrm{GL}_2(\mathbb{F}_p)$.*

Proof. Since the characteristic polynomial of $\bar{A}$ is equal to $T^2 - 2T + 1 \in \mathbb{F}_p[T]$, $\bar{A}$ is either equal to $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ or is conjugate to $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. In particular, after conjugating A by a suitable element of $\mathrm{GL}_2(\mathbb{Z}_p)$, we may and do assume that

$$A = \begin{pmatrix} 1+pa & b \\ c & 1+pd \end{pmatrix}$$

for suitable elements $a, b, c, d \in \mathbb{Z}_p$. The assumptions on A imply that $a + d \equiv 0 \pmod{p}$ and $bc \in p\mathbb{Z}_p^\times$. The fact that $bc \in p\mathbb{Z}_p^\times$ implies that $\bar{A}$ is either an upper-triangular non-trivial unipotent matrix or a lower-triangular non-trivial unipotent matrix. □

We choose and fix a generator $\bar{b}$ of the cyclic group $\mathbb{F}_{p^2}^\times$. Denote by $\bar{f}(T)$ the characteristic polynomial of $\bar{b}$ over $\mathbb{F}_p$.

Lemma 3.3. *Let $\bar{\Delta}$ be the subgroup of all diagonal matrices in $\mathrm{GL}_2(\mathbb{F}_p)$. Let $\bar{H}$ be a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ containing $\bar{\Delta}$. Assume that $\bar{H}$ contains a non-trivial unipotent element $\bar{u}$ of $\mathrm{GL}_2(\mathbb{F}_p)$, and also an element $\bar{v} \in \mathrm{GL}_2(\mathbb{F}_p)$ such that $\bar{f}(\bar{v}) = 0$ in $\mathrm{M}_2(\mathbb{F}_p)$. Then $\bar{H} = \mathrm{GL}_2(\mathbb{F}_p)$.*

Proof. The assumptions on $\bar{H}$ implies that it intersects non-trivially with every conjugacy class of $\mathrm{GL}_2(\mathbb{F}_p)$. It is a standard exercise in group theory that if a subgroup S of a finite group G intersects non-trivially with every conjugacy class in G, then $S = G$. □

Remark 3.4.

(i) The point of Lemma 3.2 is that, by imposing congruence conditions on the characteristic polynomial of a semisimple element of $\mathrm{GL}_2(\mathbb{Q}_p)$ which belongs to $\mathrm{GL}_2(\mathbb{Z}_p)$, one can make sure that the reduction of this element is a non-trivial unipotent element of $\mathrm{GL}_2(\mathbb{F}_p)$. Of course this is a general phenomenon in the context of reductive group over local fields and not restricted to GL_2.

(ii) The author claims no novelty whatsoever about Lemmas 3.1, 3.2 and 3.3. As pointed out by the referee, similar statements already appeared in Lemma 5 and Lemma 1 of [**25**].

3.3. First proof of Theorem 2.1 As explained in § 3.2, it suffices to prove the case when $g = 2$. Choose an identification of $\mathrm{GL}(\mathrm{T}_0)$ with $\mathrm{GL}_2(\mathbb{Z}_p)$. Moreover, Igusa's theorem implies that the image of the p-adic monodromy homomorphism ρ_p contains a conjugate of the subgroup Δ of diagonal matrices in $\mathrm{GL}_2(\mathbb{Z}_p)$: Consider the standard embedding $i : \mathcal{A}_{1,n}^{\mathrm{ord}} \times_{\mathrm{Spec}(\overline{\mathbb{F}}_p)} \mathcal{A}_{1,n}^{\mathrm{ord}} \to \mathcal{A}_{2,n}^{\mathrm{ord}}$. The image of the restriction of ρ_p to $\pi_1(\mathcal{A}_{1,n}^{\mathrm{ord}} \times_{\mathrm{Spec}(\overline{\mathbb{F}}_p)} \mathcal{A}_{1,n}^{\mathrm{ord}}, x_0)$ is (a conjugate of) Δ by Igusa and the functoriality of π_1.

Let H be the image of ρ_p, and let $\bar{H}$ be the image of H in $\mathrm{GL}_2(\mathbb{F}_p)$. By Lemmas 3.1 and 3.3, it suffices to show that $\bar{H}$ contains a non-trivial unipotent element $\bar{u} \in \mathrm{GL}_2(\mathbb{F}_p)$ and also an element $\bar{v}$ such that $\bar{f}(\bar{v}) = 0$ in $\mathrm{M}_2(\mathbb{F}_p)$.

Choose a quadratic polynomial $f(T) \in \mathbb{Z}[T]$ such that $f(T)$ splits over $\mathbb{R}$ and $f(T) \equiv \bar{f}(T) \pmod{p}$. Choose a quadratic polynomial $g(T) \in \mathbb{Z}[T]$ such that $g(T)$ splits over $\mathbb{R}$ and that $g(T) \equiv T^2 - 2T + 1 + p \pmod{p^2}$. Let $F_f = \mathbb{Q}[T]/(f(T))$ and let $F_g = \mathbb{Q}[T]/(g(T))$; they are the real quadratic fields defined by $f(T)$ and $g(T)$ respectively. Let a_f be the image of T in F_f and let a_g be the image of T in F_g. Choose a positive integer n_1 such that $(n_1, np) = 1$ and every element of the strict ideal class group of F_f is represented by an ideal of $\mathcal{O}_{F_f}$ which divides $n_1\mathcal{O}_{F_f}$. Similarly, choose a positive integer n_2 such that $(n_2, np) = 1$ and every element of the strict ideal class group of F_g is represented by an ideal of $\mathcal{O}_{F_g}$ which divides $n_2\mathcal{O}_{F_g}$. Choose a suitable power $q = p^r$ of p such that $\mathcal{A}_{2,n}^{\mathrm{ord}}$, $\mathcal{M}_{F_f,n}^{\mathrm{ord}}$, $\mathcal{M}_{F_g,n}^{\mathrm{ord}}$ are all defined over $\mathbb{F}_q$. Here $\mathcal{M}_{F_f,n}^{\mathrm{ord}}$ (respectively, $\mathcal{M}_{F_g,n}^{\mathrm{ord}}$) denotes the ordinary locus of the Hilbert modular surface with level-n structure attached to the real quadratic field F_f (respectively, F_g). The p-adic monodromy representation ρ_p attached to the modular variety $\mathcal{A}_{2,n}^{\mathrm{ord}}$ over $\overline{\mathbb{F}}_p$ extends to a homomorphism ρ_p^{arith} from the arithmetic fundamental group $\pi_1(\mathcal{A}_{2,n/\mathbb{F}_q}^{\mathrm{ord}}, x)$ to $\mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_2(\mathbb{Z}_p)$. Analogous statements hold for the Hilbert modular surfaces $\mathcal{M}_{F_f,n}^{\mathrm{ord}}$ and $\mathcal{M}_{F_g,n}^{\mathrm{ord}}$. Replacing r by a suitable multiple if necessary, we may and do assume that $q \equiv 1 \pmod{n^2 n_3^2}$, where n_3 is the least common multiple of n_1 and n_2.

Apply the argument in [**10**,**24**]: choose an element $b_1 \in \mathcal{O}_{F_f}$ such that $b_1 \equiv a_f \pmod{p}$ and $b_1 \equiv 2 \pmod{n^2 n_1^2}$. Also, choose an element $b_2 \in \mathcal{O}_{F_f}$ such that $b_2 \equiv 1 \pmod{p}$ and $b_2 \equiv 2 \pmod{n^2 n_1^2}$. For s sufficiently large, the quadratic polynomials $T^2 - b_1 T + q^s \in F_f[T]$ and $T^2 - b_2 T + q^s \in F_f[T]$ define q^s-Weil numbers π_1, π_2 in totally imaginary quadratic extensions K_1, K_2 of F_f such that $K_i \otimes_{\mathbb{Q}} \mathbb{Q}_p \cong (F_f \otimes_{\mathbb{Q}} \mathbb{Q}_p) \times (F_f \otimes_{\mathbb{Q}} \mathbb{Q}_p)$ for $i = 1, 2$. Moreover, the image of π_i in exactly one of the two factors is congruent to b_i modulo p. The congruence condition modulo $n^2 n_1^2$ guarantees that $(\pi_i - 1)/(n n_1)$ is integral, i.e. it is an element of $\mathcal{O}_{K_i}$, $i = 1, 2$. Therefore, we obtain abelian varieties A_1, A_2 over $\mathbb{F}_{q^s}$ with endomorphism by $\mathcal{O}_{F_f}$. The finite étale group schemes $A_i[n]$ over $\mathbb{F}_{q^s}$ is constant because $(\pi_i - 1)/n$ is integral. Changing A_i by a suitable $\mathbb{F}_{q^s}$-rational $\mathcal{O}_{F_f}$-linear isogeny, we may assume that the polarization sheaf of A_i is trivial for $i = 1, 2$. So we obtain $\mathbb{F}_{q^s}$-points of $\mathcal{A}^{\mathrm{ord}}_{2,n/\mathbb{F}_q}$. The difference $\mathrm{Fr}_{A_1} \cdot \mathrm{Fr}_{A_2}^{-1}$ of their q^s-Frobenius gives an element of the geometric fundamental group of $\mathcal{A}^{\mathrm{ord}}_{2,n}$ whose image in $\mathrm{GL}_2(\mathbb{F}_p)$ is a root of the irreducible polynomial $\bar{f}(T)$ over $\mathbb{F}_p$. Hence $\bar{H}$ contains an element $\bar{v}$ such that $\bar{f}(\bar{v}) = 0$ in $\mathrm{M}_2(\mathbb{F}_p)$.

A similar argument shows that the image of ρ_p contains an element of $\mathrm{GL}_2(\mathbb{Z}_p)$ whose characteristic polynomial is congruent to $g(T)$ modulo p^2. So $\bar{H}$ contains a non-trivial unipotent element by Lemma 3.2. We conclude that $H = \mathrm{GL}_2(\mathbb{Z}_p)$ by Lemma 3.3. $\square$

Remark 3.5. At the end of the second-to-last paragraph of the proof in § 3.3, the statement that the 'difference' of two Frobenius elements at closed points gives an element of the geometric fundamental group is a consequence of the following general fact.

Let G_1, G_2 be Barsotti–Tate groups over $\mathbb{F}_q$, and consider the natural action of $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_q)$ on the set

$$I := \mathrm{Isom}_{\overline{\mathbb{F}}_p}(G_1 \times_{\mathrm{Spec}(\mathbb{F}_q)} \times \mathrm{Spec}(\overline{\mathbb{F}}_p), G_2 \times_{\mathrm{Spec}(\mathbb{F}_q)} \times \mathrm{Spec}(\overline{\mathbb{F}}_p)).$$

Then the action of the arithmetic Frobenius element $\phi = \phi_q$ on I is given by $\alpha \mapsto \mathrm{Fr}_{G_2/\mathbb{F}_q} \circ \alpha \circ \mathrm{Fr}_{G_1/\mathbb{F}_q}^{-1}$ for all $\alpha \in I$.

Notice that in the above formula the partial compositions $\alpha \mapsto \mathrm{Fr}_{G_2/\mathbb{F}_q} \circ \alpha$ and $\alpha \circ \mathrm{Fr}_{G_1/\mathbb{F}_q}^{-1}$ are quasi-isogenies but not elements of I.

The above assertion can be seen from the following diagram

$$\begin{array}{ccc}
G_1 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p) & \xrightarrow{\beta} & G_2 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p) \\
{\scriptstyle \mathrm{Id}_{G_1} \times \sigma}\downarrow & & \downarrow{\scriptstyle \mathrm{Id}_{G_2} \times \sigma} \\
G_1 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p) & \xrightarrow{\alpha} & G_2 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p) \\
{\scriptstyle \mathrm{Fr}_{G_1} \times \mathrm{Id}_{\overline{\mathbb{F}}_p}}\downarrow & & \downarrow{\scriptstyle \mathrm{Fr}_{G_1} \times \mathrm{Id}_{\overline{\mathbb{F}}_p}} \\
G_1 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p) & \xrightarrow{\beta'} & G_2 \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p)
\end{array}$$

where β and β' are defined by the requirement that the diagram commutes. By definition of the Galois action on I, we have ${}^{\phi}\alpha = \beta$. On the other hand, the composition $\mathrm{Fr}_{G_i} \times \mathrm{Id}_{\overline{\mathbb{F}}_p} \circ \mathrm{Id}_{G_i} \times \sigma$ is equal to the absolute q-Frobenius morphism for $G_i \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathrm{Spec}(\overline{\mathbb{F}}_p)$. The commutativity of the diagram implies that $\beta' = \beta$. We have proved the assertion.

In the context of § 3.3, consider the sheaf $\mathcal{I} := \underline{\mathrm{Isom}}_{\mathcal{A}_{2,n}^{\mathrm{ord,perf}}}((A_{s_0}, \lambda_{s_0})[p^\infty], (A, \lambda)[p^\infty])$ over the perfection $\mathcal{A}_{2,n}^{\mathrm{ord,perf}}$ of $\mathcal{A}_{2,n}^{\mathrm{ord}}$ as in § 2.4. Let s_i be the closed point corresponding to the principally polarized abelian varieties (A_i, λ_i), $i = 1, 2$. Then we obtain two conjugacy classes $\mathrm{Fr}_{A_0}^{-1} \cdot \mathrm{Fr}_{A_i}$, $i = 1, 2$, in the image of the arithmetic fundamental group $\pi_1(\mathcal{A}_{2,n}^{\mathrm{ord}})$, both lying above the arithmetic Frobenius element ϕ_q. So the image of the geometric fundamental group $\pi_1(\mathcal{A}_{2,n}^{\mathrm{ord}} \times_{\mathrm{Spec}(\mathbb{F}_p)} \mathrm{Spec}(\overline{\mathbb{F}}_p))$ contains the conjugacy class $\mathrm{Fr}_{A_1} \cdot \mathrm{Fr}_{A_2}^{-1}$. That this 'difference' is well defined modulo p^N has been explained in § 3.1.

4. Hecke translation of Shimura subvarieties

4.1. Notation and sketch of idea

We follow the notation in § 2.2. In addition, we assume that $p > 2$. See Remark 4.2 for the case when $p = 2$.

The idea of our second proof of Theorem 2.1 is as follows. The unitary group H attached to the semisimple algebra $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathrm{M}_g(K)$ with involution gives rise to Hecke correspondences on $\mathcal{A}_{g,n}^{\mathrm{ord}}$ with x_0 as a fixed point. The product of g copies of the modular curve $\mathcal{A}_{1,n}^{\mathrm{ord}}$ is diagonally embedded in $\mathcal{A}_{g,n}^{\mathrm{ord}}$ as a subvariety B. The image $\rho_p(\pi_1(B, x_0))$ of the fundamental group of B under ρ_p is the subgroup D of diagonal matrices in $\mathrm{GL}_g(\mathbb{Z}_p)$, $D \cong (\mathbb{Z}_p^\times)^g$. Take an element $\gamma \in \mathrm{H}(\mathbb{Z}_{(p)})$. Such an element γ gives rise to a prime-to-p Hecke correspondence on $\mathcal{A}_{g,n}^{\mathrm{ord}}$ which has x_0 as a fixed point; the image of B under this Hecke correspondence is a subvariety $\gamma \cdot B$ in $\mathcal{A}_{g,n}^{\mathrm{ord}}$ such that $\rho_p(\pi_1(\gamma \cdot B, x_0))$ is equal to $\mathrm{Ad}(\gamma) \cdot \Delta$, the conjugation of D by the image of γ in $\mathrm{GL}_g(\mathbb{Z}_p)$. An exercise in group theory tells us that subgroups of the form $\mathrm{Ad}(\gamma) \cdot D$ generate $\mathrm{GL}_g(\mathbb{Z}_p)$. This proof was sketched in [**5**].

4.2. Second proof of Theorem 2.1 when $p > 2$

Let $\mathcal{B}$ be the product of g copies of $\mathcal{A}_{1,n}$, diagonally embedded in $\mathcal{A}_{g,n}$. Recall that E_1 is an ordinary elliptic curve over $\overline{\mathbb{F}}_p$, A_0 is the product of g copies of E_1, and λ_0 is the product principal polarization on A_0. We have $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p \cong \mathbb{Z}_p \times \mathbb{Z}_p$, corresponding to the natural splitting of $E_1[p^\infty]$ into the product of its multiplicative part $E_1[p^\infty]_{\mathrm{mult}}$ and its maximal étale quotient $E_1[p^\infty]_{\mathrm{et}}$. So we have an isomorphism $\mathrm{End}(A_0) \cong \mathrm{M}_g(\mathcal{O})$, and a splitting $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Z}_p \cong \mathrm{M}_g(\mathbb{Z}_p) \times \mathrm{M}_g(\mathbb{Z}_p)$ corresponding to the splitting of $A_0[p^\infty]$ into the product its multiplicative and étale parts. Denote by $\mathrm{pr} : (\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Z}_p)^\times \to \mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_g(\mathbb{Z}_p)$ the projection corresponding to the action of $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Z}_p$ on the étale factor $A_0[p^\infty]_{\mathrm{et}}$ of $A_0[p^\infty]$. The Rosati involution $*$ on $\mathrm{End}(A_0)$ interchanges the two factors of $\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Z}_p$. Recall that H denotes the unitary group attached to $(\mathrm{End}(A_0) \otimes_{\mathbb{Z}} \mathbb{Q}, *)$; in particular $\mathrm{H}(\mathbb{Z}_p)$ is a compact open subgroup

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

of $\mathrm{H}(\mathbb{Q}_p)$ isomorphic to $\mathrm{GL}(\mathrm{T}_0)$ under the projection map pr. Moreover, the image of $\mathrm{H}(\mathbb{Z}_{(p)})$ in $\mathrm{H}(\mathbb{Z}_p) \cong \mathrm{GL}(\mathrm{T}_0)$ is dense in $\mathrm{H}(\mathbb{Z}_p)$.

By Igusa's theorem in [**18**], the p-adic monodromy group of the restriction to $\mathcal{B}$, i.e. $\rho(\mathrm{Im}(\pi_1(\mathcal{B}, x_0) \to \pi_1(\mathcal{A}_{g,n}, x_0)))$, is naturally identified with the product of g copies of $\mathbb{Z}_p^\times$ diagonally embedded in $\mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_g(\mathbb{Z}_p)$. Denote by D this subgroup of $\mathrm{GL}(\mathrm{T}_0)$.

Lemma 4.1. *The image of the p-adic monodromy homomorphism*

$$\rho_p : \pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0) \to \mathrm{GL}(\mathrm{T}_0)$$

is a closed normal subgroup of $\mathrm{GL}(\mathrm{T}_0)$ *which contains the subgroup* $D \cong (\mathbb{Z}_p^\times)^g$.

Proof of Lemma 4.1. Every element $u \in \mathrm{H}(\mathbb{Z}_{(p)})$ defines a prime-to-p isogeny from A_0 to itself respecting the polarization λ_0. Such an element $u \in \mathrm{H}(\mathbb{Z}_{(p)})$ gives rise to

- a prime-to-p Hecke correspondence h on $\mathcal{A}_{g,n}$ having x_0 as a fixed point, and
- an irreducible component $\mathcal{B}'$ of the image of $\mathcal{B}$ under h such that $\mathcal{B}' \ni x_0$.

By the functoriality of the fundamental group, the image of the fundamental group $\pi_1(\mathcal{B}', x_0)$ of $\mathcal{B}'$ in $\pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0)$ is mapped under the p-adic monodromy representation ρ to the conjugation of D by the element $\mathrm{pr}(h) \in \mathrm{GL}(\mathrm{T}_0)$. In particular, $\rho(\pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0))$ is a closed subgroup of $\mathrm{GL}(\mathrm{T}_0)$ which contains all conjugates of D by elements in the image of $\mathrm{pr} : \mathrm{H}(\mathbb{Z}_{(p)}) \to \mathrm{H}(\mathbb{Z}_p) \cong \mathrm{GL}(\mathrm{T}_0)$.

Recall that the image of $\mathrm{H}(\mathbb{Z}_{(p)})$ in $\mathrm{H}(\mathbb{Z}_p)$ is a dense subgroup. So $\rho(\pi_1(\mathcal{A}_{g,n}^{\mathrm{ord}}, x_0))$ is a closed normal subgroup of $\mathrm{GL}(\mathrm{T}_0) \cong \mathrm{GL}_g(\mathbb{Z}_p)$ which contains the subgroup D of all diagonal elements. □

4.3. End of the second proof

An exercise in group theory shows that the only closed normal subgroup which contains the subgroup of all diagonal matrices in $\mathrm{GL}_g(\mathbb{Z}_p)$ is $\mathrm{GL}_g(\mathbb{Z}_p)$ itself. Let N be such a normal subgroup. Then N contains all matrices of the form $h \cdot u \cdot h^{-1} \cdot u^{-1} = (\mathrm{Ad}(h) \cdot u) \cdot u^{-1}$, where $h \in D$ is a diagonal matrix in $\mathrm{GL}_g(\mathbb{Z}_p)$ and u is an upper triangular unipotent matrix in $\mathrm{GL}_g(\mathbb{Z}_p)$. Since $p > 2$, not every element of $\mathbb{Z}_p^\times$ is congruent to 1 modulo p, therefore N contains all upper triangular unipotent matrices in $\mathrm{GL}_g(\mathbb{Z}_p)$. Similarly N contains all lower triangular unipotent matrices in $\mathrm{GL}_g(\mathbb{Z}_p)$. These unipotent matrices and D generate $\mathrm{GL}_g(\mathbb{Z}_p)$. □

Remark 4.2. When $p = 2$, the smallest closed normal subgroup of $\mathrm{GL}_g(\mathbb{Z}_2)$ which contains the group D of all diagonal matrices in $\mathrm{GL}_g(\mathbb{Z}_2)$ is the principal congruence subgroup U_1 of $\mathrm{GL}_g(\mathbb{Z}_2)$ of level 1, i.e. the subgroup consisting of all matrices in $\mathrm{M}_g(\mathbb{Z}_2)$ which are congruent to Id_g modulo 2. In other words, the 2-adic monodromy generated by the fundamental group of the Hilbert modular variety attached to $E = \mathbb{Q} \times \cdots \times \mathbb{Q}$ and its Hecke translates by the stabilizer at x_0 is equal to the principal congruence subgroup U_1 of $\mathrm{GL}_g(\mathbb{Z}_2)$. One way to get the full target group $\mathrm{GL}_g(\mathbb{Z}_2)$ is to use Hecke translates of the Hilbert modular variety attached to a totally real number field F with $[F : \mathbb{Q}] = g$ which is not totally split above p.

Remark 4.3. In the proof of the key Lemma 4.1 it is important to know the image of the fundamental group $\pi_1(\mathcal{B}', x_0)$ under the p-adic monodromy representation ρ_p 'on the nose'; knowing it only up to conjugation is useless. One function of the chosen base point x_0 is to help identifying the subgroup $\rho_p(\pi_1(\mathcal{B}', x_0))$ of $\mathrm{GL}_g(\mathbb{Z}_p)$; this is possible because $\mathcal{B}'$ passes through x_0. The crucial property of A_0 that $\mathrm{End}(A_0)$ is large allows us to construct many subvarieties of the form $\mathcal{B}'$. The fact that $\mathrm{End}(A_0)$ is so large that the $\mathbb{Z}_p$-points of the unitary group attached to $(\mathrm{End}(A_0), *)$ is already isomorphic to the target $\mathrm{GL}_g(\mathbb{Z}_p)$ of the homomorphism ρ has the following consequence. Subgroups of the form $\rho_p(\pi_1(\mathcal{B}', x_0))$ generate a normal subgroup of $\mathrm{GL}_g(\mathbb{Z}_p)$.

5. Hecke correspondence and p-adic monodromy

5.1. Notation and sketch of idea

We keep the notation that was used in § 2.2. Fix a positive integer $m > 0$. Consider the finite étale cover $\pi_{n;m} : \mathcal{A}_{g,n;m} \to \mathcal{A}_{g,n}^{\mathrm{ord}}$ over $\overline{\mathbb{F}}_p$ with Galois group $\mathrm{GL}_g(\mathbb{Z}/p^m\mathbb{Z})$, where $\mathcal{A}_{g,n;m}$ is the moduli space which classifies ordinary g-dimensional principally polarized ordinary abelian varieties (A, λ) with a level-n structure, plus an isomorphism $\psi : (\mathbb{Z}/p^m\mathbb{Z})^g \xrightarrow{\sim} A[p^m]_{\mathrm{et}}$, and the map π_m sends (A, λ, η, ψ) to (A, λ, η). We have the prime-to-p towers $\tilde{\mathcal{A}}_{g,n(p);m} := (\mathcal{A}_{g,nb;m})_{b \in I}$ and $\tilde{\mathcal{A}}_{g,n(p)}^{\mathrm{ord}} := (\mathcal{A}_{g,nb}^{\mathrm{ord}})_{b \in I}$, where the indexing set I consists of all positive integers such that $(b, p) = 1$. Moreover, we have a natural morphism π_m from the tower $\tilde{\mathcal{A}}_{g,n(p);m}$ to the tower $\tilde{\mathcal{A}}_{g,n(p)}^{\mathrm{ord}}$ which is compatible with $\pi_{n;m}$. The group $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ operates on the two towers, and π_m is $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$-equivariant. The $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$-action induces prime-to-p Hecke correspondences on $\mathcal{A}_{g,n;m}$ and $\mathcal{A}_{g,n}^{\mathrm{ord}}$, and the finite étale morphism $\pi_{m;n}$ is Hecke equivariant.

The statement of Theorem 2.1 is equivalent to the assertion that $\mathcal{A}_{g,n;m}$ is irreducible for every $m > 0$, since $\mathcal{A}_{g,n}^{\mathrm{ord}}$ is known to be irreducible. It suffices to show that any two points $y_1, y_2 \in \mathcal{A}_{g,n;m}$ above the base point x_0 belong to the same irreducible component of $\mathcal{A}_{g,n;m}$. Choose an element $h \in \mathrm{H}(\mathbb{Z}_{(p)}) = \mathrm{H}(\mathbb{Q}) \cap \mathrm{H}(\mathbb{Z}_p)$ such that $h_{p,\mathrm{et}} \circ \psi_1 = \psi_2$. Here ψ_1 (respectively, ψ_2) is the isomorphism $(\mathbb{Z}/p^m\mathbb{Z})^g \xrightarrow{\sim} A_0[p^m]_{\mathrm{et}}$ attached to y_1 (respectively, y_2), and $h_{p,\mathrm{et}}$ is the automorphism of $A_0[p^m]_{\mathrm{et}}$ induced by the element $h \in \mathrm{H}(\mathbb{Z}_{(p)})$. Such an element h exists because $\mathrm{H}(\mathbb{Z}_{(p)})$ is dense in $\mathrm{H}(\mathbb{Z}_p)$ for the p-adic topology.

Let $h^{(p)} \in \prod_{\ell \neq p} \mathrm{H}(\mathbb{Q}_\ell)$ be the finite prime-to-p component of h. Then y_1 belongs to the image of y_2 under the prime-to-p Hecke correspondence given by $h^{(p)}$. Now one can apply (the argument of) the main result of [**4**] to conclude that y_1 and y_2 lie on the same irreducible component of $\mathcal{A}_{g,n;m}$. This finishes the sketch of the third proof of Theorem 2.1. The actual proof consists of Lemma 5.1 and Proposition 5.2: it is clear that together they imply Theorem 2.1.

5.2. We recall the notion of abelian varieties up to prime-to-p isogenies. Denote by $\mathfrak{AV}_k$ the category of abelian varieties over k such that morphisms are homomorphisms of abelian varieties. Recall that an isogeny $\alpha : A \to B$ between abelian varieties is said to be *prime-to-p* if $\mathrm{Ker}(\alpha)$ is killed by an integer N not divisible by p. Denote by $\mathfrak{AV}_k^{(p)}$ the category of abelian varieties over k up to prime-to-p isogenies, obtained by $\mathfrak{AV}_k$

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

by formally inverting all prime-to-p isogenies. The latter category has the same objects but more morphisms: Let A, B be abelian varieties over k. Then $\mathrm{Hom}_{\mathfrak{AV}_k^{(p)}}([A],[B]) := \mathrm{Hom}_k(A,B)\otimes_{\mathbb{Z}}\mathbb{Z}_{(p)}$. Composition of morphisms in $\mathfrak{AV}_k^{(p)}$ is defined in the obvious way. In the above $[A]$ (respectively, $[B]$) denotes the image of A (respectively, B) under the obvious functor $\pi : \mathfrak{AV}_k \to \mathfrak{AV}_k^{(p)}$. We will use this notation when we want to consider an abelian variety as an object in $\mathfrak{AV}_k^{(p)}$.

An alternative definition of morphisms in $\mathfrak{AV}_k^{(p)}$ is as follows. A morphism from A to B in $\mathfrak{AV}_k^{(p)}$ is a diagram of the form $\beta_1\alpha_1^{-1} : A \xleftarrow{\alpha_1} A_1 \xrightarrow{\beta_1} B$ in $\mathfrak{AV}_k$, where α_1 is a prime-to-p isogeny and β_1 is a homomorphism from A_1 to B. Two arrows $\beta_1\alpha_1^{-1} : A \xleftarrow{\alpha_1} A_1 \xrightarrow{\beta_1} B$ and $\beta_2\alpha_2^{-1} : A \xleftarrow{\alpha_2} A_2 \xrightarrow{\beta_2} B$ are equal if and only if there exists prime-to-p isogenies $A_3 \xrightarrow{\gamma_1} A_1$, $A_3 \xrightarrow{\gamma_2} A_2$ such that $\beta_1\circ\gamma_1 = \beta_2\circ\gamma_2$. Composition of arrows in $\mathfrak{AV}_k^{(p)}$ is defined, because for every diagram $A_1 \xrightarrow{\beta_1} B \xleftarrow{\gamma_1} B_1$ in $\mathfrak{AV}_k$ where γ_1 is a prime-to-p isogeny, there exists a diagram $A_1 \xleftarrow{\beta_2} A_2 \xrightarrow{\gamma_2} B_1$ such that $\beta_1\circ\beta_2 = \gamma_1\circ\gamma_2$.

5.3. Hecke correspondence on $\mathcal{A}_{g,n;m}$

We explain the action of the group $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ on the tower $\tilde{\mathcal{A}}_{g,n(p);m}$; the action of $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ on $\tilde{\mathcal{A}}_{g,n(p)}^{\mathrm{ord}}$ is similar but simpler. See §§ 5, 6 of [**20**] for further discussion of prime-to-p Hecke correspondences on the prime-to-p tower of modular varieties of PEL-type. It is clear that the projective system $\tilde{\mathcal{A}}_{g,n(p);m}$ is isomorphic to the projective system $\tilde{\mathcal{A}}_{g,(p);m} := (\mathcal{A}_{g,N;m})_{N\in I'}$, where the indexing set I' consists of all integers N such that $N \geqslant 3$ and $(N,p)=1$. We will describe geometric points of the tower $\tilde{\mathcal{A}}_{g,(p);m}$ and the action of $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ on geometric points of $\tilde{\mathcal{A}}_{g,(p);m}$.

Let $k \supseteq \overline{\mathbb{F}}_p$ be an algebraically closed field. We fix an isomorphism $\chi : \hat{\mathbb{Z}}^{(p)} \xrightarrow{\sim} \hat{\mathbb{Z}}^{(p)}(1)$ over $\overline{\mathbb{F}}_p$. Also we fix a 'standard symplectic pairing' $\langle\cdot,\cdot\rangle : (\hat{\mathbb{Z}}^{(p)})^{2g}\times(\hat{\mathbb{Z}}^{(p)})^{2g} \to \hat{\mathbb{Z}}^{(p)}$ given by the formula $\langle(v_1,v_2),(w_1,w_2)\rangle = {}^{\mathrm{t}}v_1\cdot w_2 - {}^{\mathrm{t}}v_2\cdot w_1$, for $v_1,v_2,w_1,w_2 \in (\hat{\mathbb{Z}}^{(p)})^g$. The set $\tilde{\mathcal{A}}_{g,(p);m}(k)$ of k-points of the pro-$\overline{\mathbb{F}}_p$-scheme $\tilde{\mathcal{A}}_{g,(p);m}$ is the set of isomorphism classes of quadruples $(A,\lambda,\eta^{(p)},\psi)$, where (A,λ) is a principally polarized g-dimensional ordinary abelian variety,

$$\eta^{(p)} : (\mathbb{A}_f^{(p)}/\mathbb{Z}_p)^{2g} = \prod_{\ell\neq p}(\mathbb{Q}_\ell/\mathbb{Z}_\ell)^{2g} \xrightarrow{\sim} \prod_{\ell\neq p} A[\ell^\infty] =: A[\text{non-}p]$$

is a symplectic isomorphism, and $\psi : (\mathbb{Z}/p^m\mathbb{Z})^g \xrightarrow{\sim} A[p^m]_{\mathrm{et}}$ is an isomorphism. In the above, the principal polarization λ induces a compatible system of pairings

$$A[N]\times A[N] \to (\mathbb{Z}/N\mathbb{Z})(1) \xrightarrow{\chi^{-1}} \mathbb{Z}/N\mathbb{Z}, \quad (N,p)=1,$$

while the standard pairing $\langle\cdot,\cdot\rangle$ on $(\hat{\mathbb{Z}}^{(p)})^{2g}$ induces a compatible system of pairings

$$\langle\cdot,\cdot\rangle_N : (N^{-1}\mathbb{Z}/\mathbb{Z})^{2g}\times(N^{-1}\mathbb{Z}/\mathbb{Z})^{2g} \to \mathbb{Z}/N\mathbb{Z},$$

in the sense that

$$\langle N_2 v, N_2 w\rangle_{N_1} \equiv \langle v,w\rangle_{N_1N_2} \pmod{N_1}$$

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

for all $v,w \in ((N_2N_1)^{-1}\mathbb{Z}/\mathbb{Z})^{2g}$. Identifying $A[N]$ with $\mathrm{H}_1(A,\mathbb{Z}/N\mathbb{N})$ for $(N,p)=1$, a prime-to-p level structure $\eta^{(p)}$ as above gives rise to symplectic isomorphisms

$$\hat{\eta} : (\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\mathbb{Z}^{(p)})$$

and

$$\tilde{\eta} : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\mathbb{A}_f^{(p)}).$$

We will use a slightly different description of the set of geometric points of $\tilde{\mathcal{A}}_{g,(p);m}$. Let $k = k^{\mathrm{alg}} \supseteq \overline{\mathbb{F}}_p$ be as above. Then the set of k-points of $\tilde{\mathcal{A}}_{g,(p),m}$ is in natural bijection with the set of all isomorphism classes of quadruples $([A],\tilde{\lambda},\tilde{\eta},\psi)$, where $[A]$ is a g-dimensional ordinary abelian variety regarded as an object of $\mathfrak{AV}_k^{(p)}$, $\tilde{\lambda}$ is a principal polarization on $[A]$ (as an object $\mathfrak{AV}_k^{(p)}$, i.e. a separable polarization on A as an object in $\mathfrak{AV}_k$), $\tilde{\eta} : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\mathbb{A}_f^{(p)})$ is a symplectic isomorphism, and $\psi : (\mathbb{Z}/p^m\mathbb{Z}) \xrightarrow{\sim} A[p^m]_{\mathrm{et}}$ is an isomorphism.

In the second description of $\tilde{\mathcal{A}}_{g,(p),m}(k)$ above, more symplectic isomorphisms $\tilde{\eta}$ are allowed, because $\tilde{\eta}$ is not required to come from an isomorphism

$$\hat{\eta} : (\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\hat{\mathbb{Z}}^{(p)});$$

this is compensated by the fact that there are more isomorphisms in the category $\mathfrak{AV}_k^{(p)}$ than in the category $\mathfrak{AV}_k$. An isomorphism from $([A_1],\tilde{\lambda}_1,\tilde{\eta}_1,\psi_1)$ to $([A_2],\tilde{\lambda}_2,\tilde{\eta}_2,\psi_2)$ is a prime-to-p isogeny $\alpha : A_1 \to A_2$ such that $\alpha\circ\tilde{\eta}_1 = \tilde{\eta}_2$, $\alpha\circ\psi_1 = \psi_2$ and $\alpha^t\circ\tilde{\lambda}_2\circ\alpha = \tilde{\lambda}_1$.

We indicate how to go from the second description of $\tilde{\mathcal{A}}_{g,(p);m}(k)$ to the first description of $\tilde{\mathcal{A}}_{g,(p);m}(k)$. Let $([A],\tilde{\lambda},\tilde{\eta},\psi)$ be a quadruple as in the previous paragraph. Then there exists an abelian variety B over k and a prime-to-p-isogeny $\alpha : B \to A$, both defined up to unique isomorphisms, such that $\alpha^{-1}\tilde{\eta} : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(B,\mathbb{A}_f^{(p)})$ induces an isomorphism $\hat{\eta} : (\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(B,\hat{\mathbb{Z}}^{(p)})$ and an isomorphism $\eta^{(p)} : (\mathbb{A}_f^{(p)}/\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} B[\text{non-}p]$. Moreover, $\lambda_B := \alpha^t\circ\tilde{\lambda}\circ\alpha$ is a principal polarization of B as an abelian variety. Let $\psi_B = \alpha^{-1}\circ\psi$. Then the quadruple $(B,\lambda_B,\eta^{(p)},\psi_B)$ is well defined up to unique isomorphism and gives a point of $\tilde{\mathcal{A}}_{g,(p);m}(k)$ according to our first description.

The right action of $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ on $\tilde{\mathcal{A}}_{g,(p);m}(k)$ is quite easy to describe in terms of the second description: an element $\gamma \in \mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$ sends a point $[([A],\tilde{\lambda},\tilde{\eta},\psi)] \in \tilde{\mathcal{A}}_{g,(p);m}(k)$ to the point $[([A],\tilde{\lambda},\tilde{\eta}\circ\gamma,\psi)]$. Things are a bit more complicated with the first description. Let $[(A,\lambda,\eta^{(p)},\psi)]$ be a point of $\tilde{\mathcal{A}}_{g,(p);m}(k)$ according to the first description. Let $\hat{\eta} : (\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\hat{\mathbb{Z}}^{(p)})$ and $\tilde{\eta} : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A,\mathbb{A}_f^{(p)})$ be the symplectic isomorphisms attached to $\bar{\eta}$. There exists an abelian variety B over k and a prime-to-p isogeny $\alpha : B \to A$, unique up to unique isomorphisms, such that

$$\alpha^{-1}\circ\bar{\eta}\circ\gamma : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(B,\mathbb{A}_f^{(p)})$$

induces symplectic isomorphisms

$$\hat{\eta}_B : (\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(B,\hat{\mathbb{Z}}^{(p)}) \quad \text{and} \quad \eta_B^{(p)} : (\mathbb{A}_f^{(p)}/\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} B[\text{non-}p].$$

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

Then $\lambda_B := \alpha^t \circ \lambda \circ \alpha$ is a principal polarization of B. Let $\psi_B := \alpha^{-1} \circ \psi$. Then $[(B, \lambda_B, \eta_B^{(p)}, \psi_B)]$ is well defined, and is the image of $[(A, \lambda, \eta^{(p)}, \psi)]$ under γ.

Lemma 5.1. *Notation as above. Recall that* $x_0 = [(A_0, \lambda_0, \eta_0)] \in \mathcal{A}_{g,n}^{\mathrm{ord}}(\overline{\mathbb{F}}_p)$, $A_0 = E_1 \times \cdots \times E_1$, λ_0 *is the product principal polarization on* A_0, *and* η_0 *is a level-*n *structure on* A_0. *Let* $y_1 = [(A_0, \lambda_0, \eta_0, \psi_1)]$, $y_2 = [(A_0, \lambda_0, \eta_0, \psi_2)]$ *be two* $\overline{\mathbb{F}}_p$*-points of* $\pi_m^{-1}(x_0)$. *Let* h *be an element of* $\mathrm{H}(\mathbb{Z}_{(p)})$ *such that* $h_{p,\mathrm{et}} \circ \psi_1 = \psi_2$, *where* $h_{p,\mathrm{et}}$ *denotes the automorphism of* $A_0[p^m]_{\mathrm{et}}$ *induced by* h. *Then* y_2 *belongs to the* $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$*-Hecke orbit of* y_1 *on the* $\mathrm{GL}_g(\mathbb{Z}/p^m\mathbb{Z})$*-cover* $\mathcal{A}_{g,n;m}$ *of* $\mathcal{A}_{g,n}^{\mathrm{ord}}$.

Proof. The element $h \in \mathrm{H}(\mathbb{Z}_{(p)})$ is a prime-to-p isogeny from A_0 to itself which respects the polarization λ_0. Let $\eta^{(p)} : (\mathbb{A}_f^{(p)}/\hat{\mathbb{Z}}^{(p)})^{2g} \xrightarrow{\sim} A_0[\text{non-}p]$ be a symplectic isomorphism which extends η_0. Let $\tilde{\eta} : (\mathbb{A}_f^{(p)})^{2g} \xrightarrow{\sim} \mathrm{H}_1(A_0, \mathbb{A}_f^{(p)})$ be the symplectic isomorphism attached to $\eta^{(p)}$. The $\overline{\mathbb{F}}_p$-point $[([A_0], \lambda_0, \tilde{\eta}, \psi_i)]$ of the tower $\tilde{\mathcal{A}}_{g,n(p);m}$ lies above the $\overline{\mathbb{F}}_p$-point y_i of $\mathcal{A}_{g,n;m}$, $i = 1, 2$. Here we have followed the notation in § 5.3, and $[A_0]$ is the object in $\mathfrak{AV}_k^{(p)}$ attached to A_0.

By definition, the prime-to-p isogeny h induces an isomorphism from $([A_0], \lambda_0, \tilde{\eta}, \psi_1)$ to $([A_0], \lambda_0, h^{(p)} \circ \tilde{\eta}, \psi_2)$. Since $h^{(p)} \circ \tilde{\eta} = \tilde{\eta} \cdot (\tilde{\eta}^{-1} \cdot h^{(p)} \cdot \tilde{\eta})$, we see from the definition of prime-to-p Hecke correspondences on $\mathcal{A}_{g,n;m}$ that y_1 belongs to the image of y_2 under the prime-to-p Hecke correspondence induced by the element $\tilde{\eta}^{-1} \cdot h^{(p)} \cdot \tilde{\eta} \in \mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$. □

Remark. A prominent feature of the above argument is the similarity to the product formula: if one changes the prime-to-p level structure $\tilde{\eta}$ by the prime-to-p component $h^{(p)}$ of a 'rational element' $h \in \mathrm{H}(\mathbb{Z}_{(p)})$ and the p-power level structure ψ by the p-adic component h_p of h, one gets back to the same point of $\mathcal{A}_{g,(p);m}$.

Proposition 5.2. *Notation as in § 5.1. Let* z_1, z_2 *be two* $\overline{\mathbb{F}}_p$*-points of* $\mathcal{A}_{g,n;m}$ *which belong to the same* $\mathrm{Sp}_{2g}(\mathbb{A}_f^{(p)})$*-Hecke orbit on* $\mathcal{A}_{g,n;m}$. *Then* z_1 *and* z_2 *belong to the same irreducible component of the smooth* $\overline{\mathbb{F}}_p$*-scheme* $\mathcal{A}_{g,n;m}$.

Notation as in Lemma 5.1. We offer two proofs: one by quoting [**4**], the other by explaining the relevant part of the argument in [**4**].

Proof A. By Proposition 4.5.4 of [**4**], z_1 and z_2 belong to the same irreducible component of the smooth $\overline{\mathbb{F}}_p$-scheme $\mathcal{A}_{g,n;m}$. We need to explain why quoting [**4**] is legitimate. In [**4**], the subvariety W is assumed to be a subscheme of $\mathcal{A}_{g,n}$, while in the present situation $\mathcal{A}_{g,n;m}$ is a finite étale cover of $\mathcal{A}_{g,n}^{\mathrm{ord}}$. However one can examine the argument in [**4**] and convince oneself that the same proof works in the present situation. □

Proof B. Let $\ell_1, \dots, \ell_r$ be distinct prime numbers, all different from p, such that z_1 and z_2 belong to the same G_L-Hecke orbit on $\mathcal{A}_{g,n;m}$, where G_L denotes the product group $G_L := \mathrm{Sp}_{2g}(\mathbb{Q}_{\ell_1}) \times \cdots \times \mathrm{Sp}_{2g}(\mathbb{A}_{\ell_r})$. Let $L = \prod_{i=1}^r \ell_i$. Consider the L-adic subtower $\tilde{\mathcal{A}}_{g,nL^\infty;m} := (\mathcal{A}_{g,nL^j;m})_{j \in \mathbb{N}}$ of the tower $(\mathcal{A}_{g,nb;m})_{b \in \mathbb{N}-p\mathbb{N}}$, and let $\pi_0(\tilde{\mathcal{A}}) := \varprojlim_j \pi_0(\mathcal{A}_{g,nL^j;m})$ be the inverse limit of the set of irreducible components of $\mathcal{A}_{g,nL^j;m}$. The group G_L operates on the tower $\tilde{\mathcal{A}}_{g,nL^\infty;m}$, inducing the G_L-Hecke correspondences on $\mathcal{A}_{g,n;m}$.

Since $\mathcal{A}_{g,n;m}$ is a finite étale cover of $\mathcal{A}_{g,n}^{\mathrm{ord}}$, the image of the L-adic monodromy representation attached to the universal abelian scheme over $\mathcal{A}_{g,n;m}$ is an open subgroup of G_L, hence $\pi_0(\tilde{\mathcal{A}})$ is a finite set. The action of the group G_L on $\tilde{\mathcal{A}}$ defines a natural action of G_L on $\pi_0(\tilde{\mathcal{A}})$.

By [**4**, Lemma 3.1], every subgroup of finite index in G_L is equal to G_L itself. Therefore, G_L operates trivially on the finite set $\pi_0(\tilde{\mathcal{A}})$. So z_1 and z_2 belong to the same irreducible component of $\mathcal{A}_{g,n;m}$. □

Remark. The above proof of Theorem 2.1 was inspired by Hida's work on p-adic monodromy in [**14**].

6. Remarks and comments

Remark 6.1. Let Γ be a finite-dimensional semisimple $\mathbb{Q}$-algebra, and let $\mathcal{O}_\Gamma$ be an order of Γ. Recall that an $\mathcal{O}_\Gamma$-linear abelian variety (A,ι) over $\overline{\mathbb{F}}_p$ is Γ-hypersymmetric if

$$\mathrm{End}_{\mathcal{O}_\Gamma}(A)\otimes_{\mathbb{Z}}\mathbb{Q}_p \xrightarrow{\sim} \mathrm{End}_{\mathcal{O}_\Gamma\otimes_{\mathbb{Z}}\mathbb{Z}_p}(A[p^\infty])\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$$

(see [**6**, Definition 6.4]). Let $\mathcal{M}$ be a modular variety of PEL-type over $\overline{\mathbb{F}}_p$ with $\mathcal{O}_\Gamma$ as the ring of prescribed endomorphisms. Then there may exist a Newton stratum Z (respectively, a leaf) on $\mathcal{M}$ with no Γ-hypersymmetric point. This happens when $\mathcal{M}$ is a Hilbert modular variety $\mathcal{M}_F$ associated to a totally real number field F, $\Gamma = F$, and Z is a Newton stratum in $\mathcal{M}_F$ (or a leaf in $\mathcal{M}_F$) such that every point of Z corresponds to an $\mathcal{O}_F$-linear abelian variety with some but not all slopes equal to $\frac{1}{2}$. Then the methods in §§ 4 and 5 do not help in proving maximality of p-adic monodromy for $Z\subset\mathcal{M}_F$, while the method in § 3 does (see [**7**, § 5] and also 6.2).

6.1. We indicate how the methods in §§ 4 and 5 can be used to prove the maximality of p-adic monodromy, or equivalently the irreducibility of the Igusa tower, for the ordinary locus of a modular variety of quasisplit $\mathrm{U}(n,n)$ type. This irreducibility statement is useful for constructing p-adic L-functions for $\mathrm{GL}(n)$ (see [**13**]). It is a special case of [**14**, Corollary 8.17] (see also [**15**, § 10]). We refer to [**13**], [**14**] and [**15**] for more information on the $\mathrm{U}(n,n)$ type modular variety and related algebraic groups.

6.1.1. *Notation*

Let K be a totally imaginary extension of a totally real number field F such that p is unramified in K and every prime ideal $\wp$ in $\mathcal{O}_F$ splits in K. Let $m \geqslant 3$ be an integer relatively prime to p. The modular variety $\mathcal{M} = \mathcal{M}_{K,\mathrm{U}(n,n),m}$ over $\overline{\mathbb{F}}_p$ classifies quadruples $(A\to S,\iota,\lambda,\eta)$, where S is a scheme over $\overline{\mathbb{F}}_p$, $A\to S$ is an abelian scheme of relative dimension $2n[F:\mathbb{Q}]$, $\iota:\mathcal{O}_K\to\mathrm{End}_S(A)$ is a ring homomorphism, $\lambda: A\to A^t$ is a principal polarization such that the Rosati involution induces complex conjugation on K, and η is a level m-structure. Moreover, one requires that the Kottwitz condition in [**20**] is satisfied for the quasisplit $\mathrm{U}(n,n)$ PEL-type for K/F. Under the present assumptions on K and p, the last condition for ordinary abelian varieties can be made explicit as

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

follows. Suppose that $A \to S$ is ordinary, i.e. the Barsotti–Tate group $A[p^\infty] \to S$ is an extension of étale Barsotti–Tate group by a multiplicative Barsotti–Tate group. Then the Kottwitz condition means that the Barsotti–Tate group $A[\wp_w^\infty] \to S$ has relative dimension $n[K_w : \mathbb{Q}_p]$ and height $2n[K_w : \mathbb{Q}_p]$ for every place w of $\mathcal{O}_K$ above p.

Let $\mathcal{M}^{\mathrm{ord}}$ be the locus of $\mathcal{M}$ over which the universal abelian scheme is ordinary; it is an open dense subscheme of $\mathcal{M}$. Let $x_0 = [(A_0, \iota_0, \lambda_0, \eta_0)] \in \mathcal{M}^{\mathrm{ord}}(\overline{\mathbb{F}}_p)$ be an $\overline{\mathbb{F}}_p$-point of the ordinary locus $\mathcal{M}^{\mathrm{ord}}$. Let $*_0$ be the Rosati involution on $\mathrm{End}_{\mathcal{O}_K \otimes \mathbb{Z}_p}(A_0[p^\infty])$ attached to the polarization λ_0. The central $(\mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Z}_p)$-algebra $\mathrm{End}_{(\mathcal{O}_K \otimes \mathbb{Z}_p)}(A_0[p^\infty]_{\mathrm{mult}})$ is isomorphic to $\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})$, where v runs through all places of F above p. Similarly $\mathrm{End}_{(\mathcal{O}_K \otimes \mathbb{Z}_p)}(A_0[p^\infty]_{\mathrm{et}})$ is isomorphic to $\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})$.

6.1.2. Denote by ν the map

$$\mathrm{End}_{(\mathcal{O}_K \otimes \mathbb{Z}_p)}(A_0[p^\infty]) = \mathrm{End}_{(\mathcal{O}_K \otimes \mathbb{Z}_p)}(A_0[p^\infty]_{\mathrm{mult}}) \times \mathrm{End}_{(\mathcal{O}_K \otimes \mathbb{Z}_p)}(A_0[p^\infty]_{\mathrm{et}}) \to (\mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Z}_p)$$

which corresponds to the map

$$\left(\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})\right) \times \left(\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})\right) \to \prod_{v|p} (\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})$$

defined by

$$((B_v)_{v|p}, (C_v)_{v|p}) \mapsto \Bigl(\det_{\mathcal{O}_K \otimes \mathcal{O}_{F_v}}(B_v) \cdot \det_{\mathcal{O}_K \otimes \mathcal{O}_{F_v}}(C_v)\Bigr)_{v|p}$$

for all elements

$$((B_v)_{v|p}, (C_v)_{v|p}) \in \left(\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})\right) \times \left(\prod_{v|p} \mathrm{M}_n(\mathcal{O}_K \otimes_{\mathcal{O}_F} \mathcal{O}_{F_v})\right).$$

Denote by U the group consisting of all elements $u \in (\mathrm{End}_{\mathcal{O}_K \otimes \mathbb{Z}_p}(A_0[p^\infty]))^\times$ such that $u \cdot *_0(u) = *_0(u) \cdot u = 1$. Let L be the subgroup of U consisting of all elements $u \in U$ such that $\nu(u) = 1$. We have a product decomposition

$$L = \prod_{v|p} L_v, \quad L_v \subset (\mathrm{End}_{\mathcal{O}_K \otimes \mathcal{O}_{F_v}}(A_0[\wp_v^\infty]))^\times,$$

and L_v is isomorphic to the subgroup of $\mathrm{GL}_n(\mathcal{O}_{F_v}) \times \mathrm{GL}_n(\mathcal{O}_{F_v})$ consisting of all pairs $(u_{1,v}, u_{2,v})$ with $\det(u_{1,v}) \cdot \det(u_{2,v}) = 1$, for each place v of F above p.

With the above notation, the p-adic monodromy group for $\mathcal{M}^{\mathrm{ord}}$ is a continuous homomorphism

$$\rho_p : \pi_1(\mathcal{M}^{\mathrm{ord}}, x_0) \to L = \prod_{v|p} L_v.$$

6.1.3. To use the method in § 4 or § 5 to show that ρ_p is surjective, one needs a point x_0 of $\mathcal{M}^{\mathrm{ord}}(\overline{\mathbb{F}}_p)$ such that

$$\mathrm{End}_{\mathcal{O}_K}(A_0) \otimes_{\mathbb{Z}} \mathbb{Z}_p \xrightarrow{\sim} \mathrm{End}_{\mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Z}_p}(A_0[p^\infty]).$$

https://doi.org/10.1017/S1474748007000199 Published online by Cambridge University Press

To see the existence of such a hypersymmetric point, choose an ordinary elliptic curve E_0 over $\overline{\mathbb{F}}_p$, a free $\mathcal{O}_K$-module N of rank n, and a $\mathcal{O}_K$-valued hermitian form ψ on N which induces an sesquilinear isomorphism from N to its $\mathcal{O}_K$-dual. Let (A_0, ι_0) be the $\mathcal{O}_K$-linear abelian variety $A_0 = N \otimes_{\mathbb{Z}} E_0$, and let λ_0 be the polarization on A_0 induced by ψ and the principal polarization on E_0. Choose a level-m structure η_0 for $(A_0, \iota_0, \lambda_0)$, we get a desired hypersymmetric point $x_0 := [(A_0, \iota_0, \lambda_0, \eta_0)]$ in $\mathcal{M}^{\mathrm{ord}}$.

6.1.4. The relevant group theoretic facts are as follows.

(a) The intersection of the unitary group attached to $(\mathrm{End}(A_0), *_0)$ with the derived group of the quasi-split unitary group $\mathrm{GU}(n, n)$ for K/F has good reduction at p, and the group of $\mathbb{Z}_p$-points of this intersection is canonically isomorphic to L.

(b) The derived group of $\mathrm{GU}(n, n)$ is simply connected.

6.1.5. The methods of §§ 5 and 4 can now be applied to the present situation.

(i) The method in § 5 is directly applicable in the above setting and gives the surjectivity of the p-adic monodromy homomorphism ρ_p.

(ii) To use the method in § 4, we need a subvariety $Z \subset \mathcal{M}^{\mathrm{ord}}$ with known p-adic monodromy. Let $\mathcal{E} \to \mathcal{A}_{1,m}^{\mathrm{ord}}$ be the universal elliptic curve over the ordinary locus of the modular curve $\mathcal{A}_{1,m}$ over $\overline{\mathbb{F}}_p$. Let Z be the product of n copies of $\mathcal{M}^{\mathrm{ord}}$. Then the construction in § 6.1.3, after choosing an $\mathcal{O}_K$-basis of N, gives an embedding from Z to $\mathcal{M}^{\mathrm{ord}}$. By Igusa's theorem, the image of the restriction to $\pi_1(Z, x_0)$ of the p-adic monodromy homomorphism ρ_p is a product $\prod_{v|p} D_v$, where D_v is a subgroup of L_v for each place v of F above p. This subgroup D_v is the subgroup of all diagonal matrices if we identify D_v with a subgroup of $\mathrm{GL}_n(\mathcal{O}_{F_v}) \times \mathrm{GL}_n(\mathcal{O}_{F_v})$. If $p > 2$, then the only closed normal subgroup of L which contains $\prod_{v|p} D_v$ is L itself. So the method of § 4 implies that ρ_p is surjective when $p > 2$.

6.2. We indicate how the three methods can be used to prove maximality of p-adic monodromy attached to a leaf in $\mathcal{A}_{g,n}$. We refer to [**22**] for the notion of leaves (see also [**3**, **7**]).

6.2.1. Let $n \geqslant 3$ be a positive integer prime to p. Let $\mathcal{C}$ be a non-supersingular leaf in $\mathcal{A}_{g,n}$ over $\overline{\mathbb{F}}_p$; $\mathcal{C}$ is a locally closed subscheme in $\mathcal{A}_{g,n}$ which is smooth over $\overline{\mathbb{F}}_p$ such that $\mathcal{C}(\overline{\mathbb{F}}_p)$ is the subset of $\mathcal{A}_{g,n}(\overline{\mathbb{F}}_p)$ consisting of all elements $x = [(A_x, \lambda_x, \eta_x)] \in \mathcal{A}_{g,n}(\overline{\mathbb{F}}_p)$ such that $(A_x[p^\infty], \lambda_x[p^\infty])$ is isomorphic to $(A_0[p^\infty], \lambda_0[p^\infty])$. Here x_0 is a fixed base point in $\mathcal{C}$, and $(A_0[p^\infty], \lambda_0[p^\infty])$ is the polarized Barsotti–Tate attached to $x_0 = [(A_0, \lambda_0, \eta_0)]$.

The p-adic monodromy homomorphism attached to $\mathcal{C}$ is a continuous homomorphism

$$\rho_p = \rho_{p,\mathcal{C}} : \pi_1(\mathcal{C}, x_0) \to \mathrm{Aut}(A_0[p^\infty], \lambda[p^\infty]).$$

6.2.2. Let $n \geqslant 3$ be a positive integer prime to p. Let F be a totally real number field, let $(\mathcal{L}, \mathcal{L}^+)$ be an invertible $\mathcal{O}_F$-module with a notion of positivity, and let $\mathcal{M}_{F,\mathcal{L},\mathcal{L}^+,n}$ be the Hilbert modular variety over $\overline{\mathbb{F}}_p$ as defined in [**9**]. Let $\mathcal{C}_F$ be a non-supersingular leaf in $\mathcal{M}_{F,\mathcal{L},\mathcal{L}^+,n}$ (see [**3**], [**29**] and [**7**, § 4]). Let $x_0 = (A_0, \iota_0, \lambda_0, \eta_0)$ be a fixed base point in $\mathcal{C}_F$, where (A_0, ι_0) is an $\mathcal{O}_F$-linear abelian variety over $\overline{\mathbb{F}}_p$ such that $\dim(A_0) = [F : \mathbb{Q}]$, λ_0 is an $(\mathcal{L}, \mathcal{L}^+)$-polarization of (A_0, ι_0) as in [**9**], and η_0 is level-n structure.

The p-adic monodromy homomorphism attached to $\mathcal{C}_F$ is a continuous homomorphism

$$\rho_p = \rho_{p,\mathcal{C}_F} : \pi_1(\mathcal{C}_F, x_0) \to \mathrm{Aut}_{\mathcal{O}_F \otimes \mathbb{Z}_p}(A_0[p^\infty], \lambda[p^\infty]).$$

6.2.3. The method in § 3 can be used to show that the p-adic monodromy homomorphism $\rho_{p,\mathcal{C}_F}$ attached to a non-supersingular leaf in a Hilbert modular variety $\mathcal{M}_{F,\mathcal{L},\mathcal{L}^+,n}$ is irreducible. See [**7**, § 4] for details.

We saw in Remark 6.1 that there may not be any F-hypersymmetric point on certain leaves $\mathcal{C}_F$ in $\mathcal{M}_{F,\mathcal{L},\mathcal{L}^+,n}$. For such a leaf the target of the homomorphism $\rho_{p,\mathcal{C}_F}$ is non-commutative, while methods in §§ 4 and 5 produce only commutative subgroups contained in the image of $\rho_{p,\mathcal{C}_F}$.

6.2.4. In [**7**, § 5] the method in § 4 was used to show that the p-adic monodromy homomorphism $\rho_{p,\mathcal{C}}$ attached to a non-supersingular leaf $\mathcal{C}$ in $\mathcal{A}_{g,n}$ is surjective; the surjectivity of $\rho_{p,\mathcal{C}_F}$ explained in § 6.2.3 supplies the necessary input data. More precisely, one first proves the surjectivity of $\rho_{p,\mathcal{C}}$ in the case when $\mathcal{C}$ is minimal, i.e. when $\mathrm{End}(A_0[p^\infty])$ is a maximal order in $\mathrm{End}(A_0[p^\infty]) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$; a hypersymmetric point x_0 is used as a base point. See [**23**] for the notion of minimal Barsotti–Tate groups. The general case is deduced from the special case when $\mathcal{C}$ is minimal.

6.2.5. Let $\mathcal{C}$ be a non-supersingular leaf in $\mathcal{A}_{g,n}$ as above. Use a hypersymmetric point x_0 in $\mathcal{C}(\overline{\mathbb{F}}_p)$, as a base point. The method in § 5 (Lemma 5.1 and Proposition 5.2) yields the surjectivity of $\rho_{p,\mathcal{C}}$ directly. In this case, the method in § 5 is more efficient than the proof in [**7**, § 5]; the latter uses the methods in §§ 3 and 4.

Acknowledgements. The author was partially supported by grant DMS04-00482 from the National Science Foundation.

It is a pleasure to thank Haruzo Hida for the email communications explaining his ideas on p-adic monodromy, one of which led to the third proof. Indeed the third proof, which happens to be a sibling of the second proof, is a characteristic-p rendering of his vision. Thanks are due to the referee, for a very careful reading, many suggestions and references. (Including the reference to [**25**] and that the irreducibility of the Igusa tower over a Shimura variety of type $\mathrm{U}(n, 1)$ was proved by P. Boyer.) The ideas of the first and the second proof occurred while the author was working on the Hecke orbit problem, and I would like to express my gratitude to Frans Oort and Chia-Fu Yu for their support, inspiration, and encouragement. I would like to thank M. Harris and R. Langlands for prompting me to understand Hida's argument, and D. Lee and E. Urban for listening to the second proof. The first and the second proof were presented during the *Spring School on Abelian Varieties* held in Utrecht and Amsterdam in May 2006;

the second and the third proof were presented in the *Workshop on Automorphic Galois Representations, L-functions and Arithmetic* held at Columbia University in June 2006. I thank the organizers of these conferences, as well as Utrecht University, the University of Amsterdam and Columbia University for their hospitality.

References

1. J. Achter and P. Norman, Local monodromy of p-divisible groups, preprint.
2. C.-L. Chai, Arithmetic minimal compactification of Hilbert–Blumenthal moduli spaces, *Ann. Math.* **131** (1990), 541–554.
3. C.-L. Chai, *Hecke orbits on Siegel modular varieties*, Progress in Mathematics, Volume 235, pp. 71–107 (Birkhäuser, 2004).
4. C.-L. Chai, Monodromy of Hecke-invariant subvarieties, *Q. J. Pure Appl. Math.* **1** (Special issue: in memory of Armand Borel) (2005), 291–303.
5. C.-L. Chai, Hecke orbits as Shimura varieties in positive characteristic, in *Proc. ICM Madrid 2006*, Volume II, pp. 295–312 (European Mathematical Society, 2006).
6. C.-L. Chai and F. Oort, Hypersymmetric abelian varieties, *Q. J. Pure Appl. Math.* **2** (Coates special issue) (2006), 1–27.
7. C.-L. Chai and F. Oort, Monodromy and irreducibility of leaves, Informal notes for a workshop on abelian varieties, Amsterdam, 29–31 May 2006.
8. P. Deligne, Variétés abéliennes ordinaires sur un corps fini, *Invent. Math.* **8** (1969), 238–243.
9. P. Deligne and G. Pappas, Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant, *Compositio Math.* **90** (1994), 59–79.
10. P. Deligne and K. Ribet, Values of abelian L-functions at negative integers over totally real fields, *Invent. Math.* **59** (1980), 227–286.
11. T. Ekedahl, The action of monodromy on torsion points of Jacobians, in *Arithmetic algebraic geoemtry* (ed. G. van der Geer, F. Oort and J. Steenbrink), Progress in Mathematics, Volume 89, pp. 41–49 (Birkhäuser, 1991).
12. G. Faltings and C.-L. Chai, *Degeneration of abelian varieties*, Ergebnisse Bd 22 (Springer, 1990).
13. M. Harris, J.-S. Li and C. M. Skinner, p-adic L functions for unitary Shimura varieties, I, Construction of the Eisenstein measure, *Documenta Math.* (Extra Volume: John H. Coates' Sixtieth Birthday, 2006), 393–464.
14. H. Hida, *p-adic automorphic forms on Shimura varieties* (Springer, 2004).
15. H. Hida, p-adic automorphic forms on reductive groups, *Astérisque* **296** (2005), 147–254.
16. H. Hida, Irreducibility of the Igusa tower, *Acta Math. Sinica*, in press.
17. T. Honda, Isogeny classes of abelian varieties over finite fields, *J. Math. Soc. Jpn* **20** (1968), 83–95.
18. J. Igusa, On the algebraic theory of elliptic modular functions, *J. Math. Soc. Jpn* **20** (1968), 96–106.
19. N. M. Katz, P-adic properties of modular schemes and modular forms, in *Modular fuctions of one variable, III*, Lecture Notes in Mathematics, Volume 350, pp. 69–190 (Springer, 1973).
20. R. E. Kottwitz, Points on some Shimura varieties over finite fields, *J. Am. Math. Soc.* **2** (1992), 373–444.
21. D. U. Lee, P-adic monodromy of the ordinary locus of Picard moduli schemes, PhD dissertation, University of Pennsylvania (2005).
22. F. Oort, Foliations in moduli spaces of abelian varieties, *J. Am. Math. Soc.* **17** (2004), 267–296.

23. F. Oort, Minimal p-divisible groups, *Ann. Math.* **161** (2005), 1021–1036.
24. K. Ribet, p-adic interpolation via Hilbert modular forms, *Proc. Symp. Pure Math.* **29** (1975), 581–592.
25. G. Shimura, A reciprocity law in non-solvable extensions, *J. Reine Angew. Math.* **221** (1966), 209–220.
26. J. Tate, Endomorphisms of abelian varieties over finite fields, *Invent. Math.* **2** (1966), 134–144.
27. J. Tate, *Classes d'isogénie de variétés abéliennes sur un corps fini (d'après T. Honda)*, Séminaires Bourbaki 21, Exp. 352 (1968/69), Lecture Notes in Mathematics, Volume 179, pp. 95–110 (Springer, 1971).
28. W. C. Waterhouse and J. S. Milne, Abelian varieties over finite fields, in *Proc. Symp. Pure Mathematics, Number Theory Institute, Stony Brook, 1969*, Volume 20, pp. 53–64 (American Mathematical Society, Providence, RI, 1971).
29. C.-F. Yu, Discrete Hecke orbit problem on Hilbert–Blumenthal modular varieties, preprint (2005).