

CORRESPONDENCES ON SHIMURA CURVES AND MAZUR'S PRINCIPLE AT p

FRAZER JARVIS

In this paper, we continue our earlier study of Mazur's Principle for totally real fields, and extend the main result to one also applicable to primes dividing the characteristic. The method we use is a naive generalisation of that of Mazur and Ribet.

1. Introduction

Let F be a totally real field of degree d over \mathbb{Q} , and let ℓ be an odd prime. We recall the main result of [12], as improved by Fujiwara's unpublished work [10]:

Theorem 1.1 (Mazur's Principle). *Assume that*

$$\bar{\rho} : \text{Gal}(\bar{F}/F) \longrightarrow \text{GL}_2(\bar{\mathbb{F}}_\ell)$$

is a continuous irreducible semisimple representation which is attached to a Hilbert cuspidal eigenform $f \in S_{k,w}(U_0(\mathfrak{p}) \cap U_1(\mathfrak{n}))$, where $\mathfrak{p} \nmid n\ell$ and $k \geq 2t$. Suppose $[F(\mu_\ell) : F] \geq 4$. Then if $\bar{\rho}$ is irreducible, and unramified at \mathfrak{p} , and $N_{F/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$, there exists a Hilbert cuspidal eigenform $f' \in S_{k,w}(U_1(\mathfrak{n}))$ to which $\bar{\rho}$ is attached.

Under an additional hypothesis, the same result is true when $[F(\mu_\ell) : F] = 2$. In [12], we proved this result with an additional hypothesis if $[F : \mathbb{Q}]$ is even; this was removed in [10]. When $[F : \mathbb{Q}]$ is odd, the results of this paper do not depend on unpublished work, but we need Fujiwara's results when $[F : \mathbb{Q}]$ is even.

Subsequently, Rajaei removed the restriction that $N_{F/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$, thus generalising Ribet's theorem, at least when $[F : \mathbb{Q}]$ is odd. These results allow one to completely lower the level away from ℓ , using other results of Fujiwara, as well as [13]. Further, Skinner and Wiles ([20]) have given a simple proof of level lowering if one allows the base field to be replaced with a suitable soluble extension.

The above theorem treats all cases with ℓ odd. If $\ell = 2$, then it never applies, as $\mathfrak{p} \nmid \ell$, which would imply that $N_{F/\mathbb{Q}}(\mathfrak{p})$ is odd, so certainly is congruent to 1 (mod 2). However, Buzzard ([3]) has a beautiful argument

which deals with mod 2 representations, and one can show ([14]) that it generalises to totally real fields.

The behaviour when $\ell = p$, the prime below \mathfrak{p} , remains less clear, however. As announced in [10], we may always assume that a modular mod p Galois representation arises from a modular form of weight $2t$ (in Hida's notation [11]). For this reason, we will focus on weight $2t$ in what follows. This allows us to work with Jacobians in a similar way to Ribet ([19]).

The main result of this paper (Theorem 6.2) is a partial result in the case $\ell = p$. In the same way as Ribet ([19]), we explain how to deduce a result along the lines of "Shimura-Taniyama-Weil implies Fermat" for totally real fields, assuming certain hypotheses on the field (which seem to be satisfied only very rarely in practice).

It would be interesting to translate the proof of the main result of this paper into the language of sheaf cohomology, and use techniques in p -adic cohomology to deduce the same result directly for arbitrary weights. However, the action of Galois on the cohomology of sheaves with p -power torsion on curves with semistable reduction mod p still seems not to be completely understood, despite impressive recent work by Breuil, Fontaine, Hyodo, Kato and Tsuji, amongst others. Indeed, our first attempts to solve this problem followed this strategy; we may return to it at a later date.

The current paper may presumably also be viewed as one of the first steps in the study of optimal weights for representations $\bar{\rho}$.

It is a pleasure to thank the referee for many useful comments.

2. Shimura curves and integral models

The proof of the above theorem, as well as the stronger version of Rajaei ([18]), is geometric in nature, involving a study of the cohomology of certain Shimura curves. This approach was motivated by the study of Carayol ([5]). We try to extend the method to give results for primes dividing p .

If v denotes a prime of F , then we let $\mathcal{O}_{F,(v)}$ denote the v -integral elements of F , and let $\mathcal{O}_{F,v}$ denote its completion, with residue field κ_v .

Let B be a quaternion algebra over F , ramified at all but one infinite place τ , and not at any places above p . We fix a ring isomorphism $\mathcal{O}_{B,v} \cong M_2(\mathcal{O}_{F,v})$ at all finite places v which split B and an isomorphism $B_\tau \cong M_2(F_\tau)$ at the unramified infinite place. The multiplicative group B^\times defines, by restriction of scalars, an algebraic group over \mathbb{Q} denoted by G .

Let \mathbb{A}^∞ denote the finite adeles (over \mathbb{Q}). Associated to any open compact subgroup K of $G(\mathbb{A}^\infty)$ is the Shimura curve

$$M_K(\mathbb{C}) = G(\mathbb{Q}) \backslash ((G(\mathbb{A}^\infty)/K) \times X).$$

Here, as with all our remarks on Shimura curves, the notation follows that of [4], so that X denotes $\mathbb{C} - \mathbb{R}$. By work of Shimura, there exists an F -scheme

M_K whose complex points are precisely $M_K(\mathbb{C})$ (where we regard F as a subfield of \mathbb{C} using the unramified infinite place τ of B).

In [5] and [12], it was necessary to study the cohomology of the special fibre of these curves, and it was thus necessary to prove the existence of models with certain properties for these curves defined over (the localisations of) the ring of integers \mathcal{O}_F of F . Although it is likely that these exist in greater generality, this has only been demonstrated for certain subgroups K . We state the main results of [4] and [12].

Let \mathfrak{p} denote a prime of F above the rational prime p . Write κ for the residue field $\kappa_{\mathfrak{p}}$. We suppose that K factors as $K_{\mathfrak{p}}H$, where H is an open compact subgroup of Γ , the restricted direct product over all finite places $v \neq \mathfrak{p}$ of $(B \otimes F_v)^\times$, as in [4], 0.4. Then we have the following results from [4] ((1) was already known to Morita):

- Theorem 2.1** (Carayol). 1) *Suppose $K_{\mathfrak{p}}$ is the subgroup $K_{\mathfrak{p}}^0 = \mathrm{GL}_2(\mathcal{O}_{F,\mathfrak{p}})$ (under the identification made above). Then, if H is sufficiently small, there exists a model $\mathbf{M}_{0,H}$ of M_K defined over $\mathcal{O}_{F,(\mathfrak{p})}$. This model is proper and smooth over $\mathrm{spec} \mathcal{O}_{F,(\mathfrak{p})}$.*
- 2) *Suppose $K_{\mathfrak{p}}$ is the subgroup $K_{\mathfrak{p}}^n$ of matrices congruent to I modulo \mathfrak{p}^n , again using the above identification. Then, if H is sufficiently small (which depends on n), there exists a regular model $\mathbf{M}_{n,H}$ of M_K with a map to $\mathbf{M}_{0,H}$. The morphism $\mathbf{M}_{n,H} \rightarrow \mathbf{M}_{0,H}$ is finite and flat.*

We will also need to think about one other case, already considered in [12]:

- Theorem 2.2.** 1) *Suppose $K_{\mathfrak{p}}$ is the subgroup*

$$U_0(\mathfrak{p}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_{F,\mathfrak{p}}) \mid c \in \mathfrak{p} \right\},$$

again using the above identification. Then, if H is sufficiently small (the same condition as 2.1(1)), there exists a regular model $\mathbf{M}_{U_0(\mathfrak{p}),H}$ of M_K defined over $\mathbf{M}_{0,H}$. The morphism $\mathbf{M}_{U_0(\mathfrak{p}),H} \rightarrow \mathbf{M}_{0,H}$ is finite and flat.

- 2) *The special fibre $\mathbf{M}_{U_0(\mathfrak{p}),H} \times \bar{\kappa}$ is isomorphic to a union of two copies of $\mathbf{M}_{0,H} \times \bar{\kappa}$ intersecting transversely above a finite set of points Σ_H .*

The set of points Σ_H are the supersingular points of $\mathbf{M}_{0,H} \times \bar{\kappa}$, and we use the same notation Σ_H for the points which lie above them in $\mathbf{M}_{U_0(\mathfrak{p}),H} \times \bar{\kappa}$, the singular points of the special fibre. We now describe this set Σ_H . Let \bar{B} denote the quaternion algebra got from B by changing the invariants at \mathfrak{p} and at τ (so it is now ramified at both these places, and is totally definite). Let \bar{G} denote the algebraic group $\mathrm{Res}_{F/\mathbb{Q}} \bar{B}^\times$, and fix, for all places $v \neq \mathfrak{p}, \tau$, an isomorphism between $B \otimes F_v$ and $\bar{B} \otimes F_v$. Then $\bar{G}(\mathbb{A}^\infty)$ may be identified

with $\Gamma \times \overline{B}_{\mathfrak{p}}^{\times}$. By [4], 11.2(3), there is a bijection

$$\begin{aligned} \Sigma_H &\cong \overline{G}(\mathbb{Q}) \backslash \overline{G}(\mathbb{A}^{\infty}) / H \times \mathcal{O}_{\overline{B}_{\mathfrak{p}}}^{\times} \\ &\cong \overline{G}(\mathbb{Q}) \backslash \Gamma \times \mathbb{F}_{\mathfrak{p}}^{\times} / H \times \mathcal{O}_{\mathbb{F}, \mathfrak{p}}^{\times} \end{aligned}$$

where the second isomorphism is induced by the reduced norm $\overline{B}_{\mathfrak{p}}^{\times} \longrightarrow \mathbb{F}_{\mathfrak{p}}^{\times}$. Then Σ_H is a finite set, and we denote its cardinality by s_H .

According to [4], 11.1.1, the action of $\mathrm{GL}_2(\mathbb{F}_{\mathfrak{p}})$ on the inverse system of Shimura curves descends to an action on the supersingular points Σ_H , and the action factors through $\det : \mathrm{GL}_2(\mathbb{F}_{\mathfrak{p}}) \longrightarrow \mathbb{F}_{\mathfrak{p}}^{\times}$ (use the second description above of Σ_H as a quotient). Further, if we normalise the reciprocity map of class field theory so that arithmetic Frobenius elements correspond to uniformisers (the opposite to [4]), then [4], 11.2(2), shows that an element $\sigma \in W(\mathbb{F}_{\mathfrak{p}}^{\mathrm{ab}}/\mathbb{F}_{\mathfrak{p}})$ acts on the set Σ_H in exactly the same way as the element $[\sigma]$ of $\mathbb{F}_{\mathfrak{p}}^{\times}$ corresponding to σ by class field theory.

We recall from [12] that above each geometric point x of $\mathbf{M}_{0,H}$ was a divisible $\mathcal{O}_{\mathfrak{p}}$ -module $\mathbf{E}_{\infty}|_x$. The study of the integral model $\mathbf{M}_{U_0(\mathfrak{p}),H}$ in characteristic p arose from classifying isogenies of divisible $\mathcal{O}_{\mathfrak{p}}$ -modules from $\mathbf{E}_{\infty}|_x$ whose kernel was a group (scheme) isomorphic to $(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})$. There were, in general, two families (which coincided at supersingular points), those arising from the Frobenius and those arising from the Verschiebung; this showed that $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}$ consisted of two copies of $\mathbf{M}_{0,H} \otimes \overline{\kappa}$ which intersected above the supersingular points. Above the point $x \in \mathbf{M}_{0,H} \otimes \overline{\kappa}$ on one copy is the point x , together with the map $F : \mathbf{E}_{\infty}|_x \longrightarrow \mathbf{E}_{\infty}|_{\sigma x}$, and on the other copy is the point σx , together with the map $V : \mathbf{E}_{\infty}|_{\sigma x} \longrightarrow \mathbf{E}_{\infty}|_x$.

3. Correspondences

We are going to define correspondences on the \mathbb{C} -points of these Shimura curves. These maps will, in fact, be defined over \mathbb{F} . These will then give rise to maps on the Jacobians, again defined over \mathbb{F} ; the Néron mapping property will then show that these maps extend to the Néron models of the Jacobians, and thus to their fibres. The only difference between this section and the discussion of correspondences in [19], is that in the setting of Shimura curves over more general totally real fields, the moduli interpretation of the correspondences used in [19] is not applicable, and instead we use an adelic approach.

We first recall that if $h \in G(\mathbb{A}^{\infty})$, and K_1 and K_2 are open compact subgroups of $G(\mathbb{A}^{\infty})$ such that $K_1 \subseteq hK_2h^{-1}$, then there is a covering map

$$\begin{aligned} h : M_{K_1}(\mathbb{C}) &\longrightarrow M_{K_2}(\mathbb{C}) \\ (g, x) &\mapsto (gh, x) \end{aligned}$$

for $g \in G(\mathbb{A}^\infty)$, $x \in X$. By Shimura's theory of canonical models, these maps come from maps on the underlying F-schemes. We will define the maps on the \mathbb{C} -points, but regard them as defined over F.

We first define the Hecke correspondence $T_{\mathfrak{p}}$. We use the same letter for a uniformiser of a prime ideal as for the ideal itself; thus \mathfrak{p} will also denote a uniformiser for the prime ideal \mathfrak{p} . Let $\pi \in G(\mathbb{A}^\infty)$ denote the element which is the identity at every place apart from \mathfrak{p} , where it is $\begin{pmatrix} 1 & 0 \\ 0 & \mathfrak{p} \end{pmatrix}$.

We define the following two degeneracy maps between the curves $M_{U_0(\mathfrak{p}^r),H}$ (where we regard $U_0(\mathfrak{p}^0)$ as $\mathrm{GL}_2(\mathcal{O}_{F,\mathfrak{p}})$):

$$\begin{aligned} 1 : M_{U_0(\mathfrak{p}^{r+1}),H} &\longrightarrow M_{U_0(\mathfrak{p}^r),H} \\ (g, x) &\mapsto (g, x) \end{aligned}$$

for $g \in G(\mathbb{A}^\infty)$, $x \in X$, and

$$\begin{aligned} \pi : M_{U_0(\mathfrak{p}^{r+1}),H} &\longrightarrow M_{U_0(\mathfrak{p}^r),H} \\ (g, x) &\mapsto (g\pi, x) \end{aligned}$$

Then we let $T_{\mathfrak{p}}$ denote the map $\pi_* 1^*$ from divisors on $M_{U_0(\mathfrak{p}^r),H}$ to themselves. This map induces an endomorphism of the Jacobian of the curve, which we denote by $J_{U_0(\mathfrak{p}^r),H}$. A lengthy, but entirely elementary, calculation shows that if $F = \mathbb{Q}$, $\mathfrak{p} = p$ and $B = \mathrm{GL}_2$, then the map T_p just defined is exactly the same map as that denoted T_p in [19], pp.443–4, the p th Hecke operator induced by Picard functoriality. (The map $1_* \pi^*$ is the Albanese version, denoted ξ_p in [19].) We have thus defined the Hecke correspondence

$$T_{\mathfrak{p}} : J_{U_0(\mathfrak{p}^r),H} \longrightarrow J_{U_0(\mathfrak{p}^r),H}.$$

In the case $r = 0$, we write $\alpha = 1$ and $\beta = \pi$.

Next we define an Atkin-Lehner automorphism $w_{\mathfrak{p}}$ of $J_{U_0(\mathfrak{p}),H}$. For this, we let $w_{\mathfrak{p}}$ denote the element of $G(\mathbb{A}^\infty)$ which is the identity at every place apart from \mathfrak{p} , where it is $\begin{pmatrix} 0 & 1 \\ -\mathfrak{p} & 0 \end{pmatrix}$. As this element of $G(\mathbb{A}^\infty)$ normalises $U_0(\mathfrak{p}) \times H$, the map $(g, x) \mapsto (gw_{\mathfrak{p}}, x)$ gives a map from $M_{U_0(\mathfrak{p}),H}$ to itself, and thus an endomorphism of $J_{U_0(\mathfrak{p}),H}$. We denote this map also by $w_{\mathfrak{p}}$.

Proposition 3.1. *Considered as maps from $J_{U_0(\mathfrak{p}),H}$ to itself, there is an equality*

$$\alpha^* \beta_* = T_{\mathfrak{p}} + w_{\mathfrak{p}}.$$

Proof. This is a calculation that will be performed entirely on \mathbb{C} -points, but as all maps are defined over F, this will give the result.

Recall that

$$\begin{aligned} M_{U_0(\mathfrak{p}^2),H} &\longrightarrow M_{U_0(\mathfrak{p}),H} \\ 1 : (g, x) &\mapsto (g, x) \\ \pi : (g, x) &\mapsto (g\pi, x) \end{aligned}$$

and $T_{\mathfrak{p}} = \pi_* 1^*$. Clearly

$$T_{\mathfrak{p}}(g, x) = \sum_h (gh\pi, x),$$

where $U_0(\mathfrak{p}) = \coprod_h hU_0(\mathfrak{p}^2)$. (Abusively, we are using h to denote an element of $U_0(\mathfrak{p})$ as well as the corresponding element of $G(\mathbb{A}^\infty)$ which is the identity at all other components.) It is easy to see that we may let h run through the set

$$\left\{ \begin{pmatrix} 1 & 0 \\ \alpha\mathfrak{p} & 1 \end{pmatrix} \right\},$$

where α runs through a set of representatives in $\mathcal{O}_{\mathbb{F},\mathfrak{p}}$ of $\mathcal{O}_{\mathbb{F},\mathfrak{p}}/\mathfrak{p}$. It follows that

$$T_{\mathfrak{p}}(g, x) = \sum \left(g \begin{pmatrix} 1 & 0 \\ \alpha\mathfrak{p} & \mathfrak{p} \end{pmatrix}, x \right).$$

Similarly,

$$\alpha^* \beta_*(g, x) = \sum (g\pi h, x),$$

where $\mathrm{GL}_2(\mathcal{O}_{\mathbb{F},\mathfrak{p}}) = \coprod_h hU_0(\mathfrak{p})$. We may let h run over the set

$$\left\{ \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\},$$

where α runs through the set of representatives in $\mathcal{O}_{\mathbb{F},\mathfrak{p}}$ of $\mathcal{O}_{\mathbb{F},\mathfrak{p}}/\mathfrak{p}$ as above. It follows that

$$\alpha^* \beta_*(g, x) = \sum \left(g \begin{pmatrix} 1 & 0 \\ \alpha\mathfrak{p} & \mathfrak{p} \end{pmatrix}, x \right) + \sum \left(g \begin{pmatrix} 0 & 1 \\ -\mathfrak{p} & 0 \end{pmatrix}, x \right).$$

Comparing the two expressions, and recalling the definition of $w_{\mathfrak{p}}$, the proposition follows. \square

If H is a subgroup which can be written as $\prod_{\mathfrak{q} \neq \mathfrak{p}} H_{\mathfrak{q}}$, then we can define correspondences $T_{\mathfrak{q}} = \begin{pmatrix} 1 & 0 \\ 0 & \mathfrak{q} \end{pmatrix}_* 1^*$ and $S_{\mathfrak{q}} = \begin{pmatrix} \mathfrak{q} & 0 \\ 0 & \mathfrak{q} \end{pmatrix}_* 1^*$ in the same way, whenever $H_{\mathfrak{q}} = \mathrm{GL}_2(\mathcal{O}_{\mathbb{F},\mathfrak{q}})$. The two maps in the first of these definitions are defined in the same way to those above; in the second, however, they are endomorphisms of the Jacobian of $M_{U_{\mathfrak{p}},H}$.

4. Hecke operators and Eisenstein ideals

Suppose that S is a finite set of places of F containing all infinite places. We will introduce an action of a Hecke algebra \mathbb{T}^S , defined as the polynomial ring over \mathbb{Z} generated by $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ for primes $\mathfrak{q} \notin S$. If $\mathfrak{p} \in S$, the Hecke algebra $\mathbb{T}^{S,\mathfrak{p}}$ will be the \mathbb{Z} -algebra defined as above but with the additional generator $T_{\mathfrak{p}}$.

Suppose S also contains all primes over which B is ramified. Then \mathbb{T}^S acts naturally, through a quotient which we will denote by $\mathbb{T}_{k,w,B}^S(U)$, on the space $S_{k,w,B}(U)$ of cusp forms of weight (k, w) for any open compact subgroup $U = \prod_{\mathfrak{q}} U_{\mathfrak{q}}$ of $G(\mathbb{A}^\infty)$ such that $U_{\mathfrak{q}} = \mathrm{GL}_2(\mathcal{O}_{F,\mathfrak{q}})$ for all $\mathfrak{q} \notin S$. (All of our notation for Hilbert modular forms follows Hida [11] and our earlier papers [12] and [13].) We say that U is an S -subgroup when this holds.

We recall from [13] the definition of Eisenstein ideals in these Hecke algebras.

If \mathfrak{m} is a maximal ideal of \mathbb{T}^S of residue characteristic p , there is a homomorphism

$$\bar{\theta}_{\mathfrak{m}} : \mathbb{T}^S \longrightarrow \mathbb{T}^S/\mathfrak{m} \hookrightarrow \overline{\mathbb{F}}_p,$$

and an eigenform whose reduction has Hecke eigenvalues given by $\bar{\theta}_{\mathfrak{m}}$; the semisimplification of the reduction of the associated Galois representation gives a representation

$$\bar{\rho}_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbb{F}}/F) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_p).$$

In fact, the representation has a model valued in the field generated by the traces; as $\mathrm{tr} \bar{\rho}_{\mathfrak{m}}(\mathrm{Frob}_{\mathfrak{q}}) = \bar{\theta}_{\mathfrak{m}}(T_{\mathfrak{q}})$ for all but finitely many \mathfrak{q} , the image of $\bar{\rho}_{\mathfrak{m}}$ may be taken to be $\mathrm{GL}_2(\mathbb{T}^S/\mathfrak{m})$.

Definition 4.1. We say that \mathfrak{m} is *Eisenstein* if and only if $\bar{\rho}_{\mathfrak{m}}$ is reducible. Equivalently, there should be some integral ideal \mathfrak{f} such that for all but finitely many primes \mathfrak{q} which are trivial in the narrow ray class group $Cl(\mathfrak{f})$, one has $T_{\mathfrak{q}} - 2 \in \mathfrak{m}$ and $S_{\mathfrak{q}} - 1 \in \mathfrak{m}$.

For more about Eisenstein ideals (and an explanation of the equivalence of the two definitions above), see [13], §3. Finally, we say that a \mathbb{T}^S -module is *Eisenstein* if all maximal ideals in its support are Eisenstein.

Similar notions exist if \mathbb{T}^S is replaced by $\mathbb{T}^{S,\mathfrak{p}}$.

5. Jacobians and Hecke operators

In this section, we briefly outline the correspondence between Jacobians and the étale cohomology of the Shimura curve M_U . We will then use the analysis of Carayol ([5]) to characterise the maximal ideals of the Hecke algebras which are in the support of the Jacobians.

Recall from [5], §2.3, that if U is open compact in $G(\mathbb{A}^\infty)$, and $E \subset \mathbb{C}$ is a number field containing the Galois closure of F and splitting B , then

$$H^1(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, E_\varphi) \otimes_{E_\varphi} \bar{E}_\varphi \cong \bigoplus_C (\pi^\infty)^U \otimes \rho_\pi^\varphi(-1),$$

where φ is a prime of E , and C denotes the set of cuspidal automorphic representations of B of weight $k = 2t$. As $(\pi^\infty)^U$ has finite dimension, and is 0 for all but finitely many $\pi \in C$, this is therefore a finite direct sum of 2-dimensional \bar{E}_φ -vector spaces. The inductive system of these isomorphisms, as U varies, is $G(\mathbb{A}^\infty)$ -equivariant.

Now suppose that U is an S -subgroup. Then the action of $G(\mathbb{A}^\infty)$ on the inductive system of cohomology groups leads, by a similar method to that outlined above for correspondences on Jacobians (see [11]), to an action of the Hecke algebra \mathbb{T}^S on $H^1(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, E_\varphi)$. The above isomorphism shows that the maximal ideals of \mathbb{T}^S in its support correspond to cuspidal automorphic representations of weight $2t$ on B with fixed vectors under U .

Choosing a $\text{Gal}(\bar{\mathbb{F}}/F)$ -stable lattice for ρ_π^φ , and taking the semisimplification $\bar{\rho}_\pi^\varphi$ of its reduction (in our application, the reduction will be irreducible, so taking the semisimplification will not be necessary), we see that $\bar{\rho}_\pi^\varphi(-1)$ will be contained in

$$(H^1(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, \mathcal{O}_{E,\varphi}) \otimes_{\mathcal{O}_{E_\varphi}} \mathcal{O}_{\bar{E}/\varphi})^{\text{ss}};$$

by the argument of [12], §15, this space embeds into

$$(H^1(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, \mathcal{O}_E/\varphi) \otimes_{\mathcal{O}_E/\varphi} \mathcal{O}_{\bar{E}/\varphi})^{\text{ss}}.$$

However, the cokernel is the ℓ -torsion in the torsion free group $H^2(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, \mathcal{O}_{E,\varphi})$, and so the embedding is actually an isomorphism. Moreover, \mathcal{O}_E/φ is flat over $\mathbb{Z}/p\mathbb{Z}$, and so this space is the same as

$$(H^1(M_U \otimes_{\mathbb{F}} \bar{\mathbb{F}}, \mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{Z}/p\mathbb{Z}} \bar{\mathbb{F}}_p)^{\text{ss}}.$$

But it is well-known that for a smooth curve over an algebraically closed field of characteristic 0, the étale cohomology with coefficients in $\mu_p = \mathbb{Z}/p\mathbb{Z}(1)$ is canonically isomorphic to the p -torsion of the Jacobian of the curve.

It follows that $\bar{\rho}_\pi^\varphi$ is a submodule of the p -torsion of the Jacobian of M_U , which we denote by J_U ; in particular, the maximal ideals of \mathbb{T}^S in the support of $J_U[p]$ contain those corresponding to cuspidal automorphic representations on B of weight $2t$ with fixed vectors under U , and conversely, any 2-dimensional irreducible submodule of (the semisimplification of) the p -torsion of J_U is the reduction of the Galois representation associated to some cuspidal automorphic representation on B of weight $2t$ with fixed vectors under U .

In the case that $U_{\mathfrak{p}} = U_0(\mathfrak{p})$, we have analogously an action of $\mathbb{T}^{S,\mathfrak{p}}$, and its quotient $\mathbb{T}_{k,w,B}^{S,\mathfrak{p}}(U)$, on the space $S_{k,w,B}(U)$, the action of the operator $T_{\mathfrak{p}}$ on the space of cusp forms being defined in the usual way ([11], p.306).

If U is an S -subgroup, it follows that there is an action of the Hecke algebra \mathbb{T}^S on J_U (using the Hecke correspondences $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ defined above) and that maximal ideals of the Hecke algebra in the support of $J_U[p]$ correspond to automorphic representations of weight $2t$ on B which have fixed vectors under U .

In the same way, if also $U_{\mathfrak{p}} = U_0(\mathfrak{p})$, the Hecke algebra $\mathbb{T}^{S,\mathfrak{p}}$ acts on J_U , and similar results hold.

6. Mazur's Principle—preliminaries

Definition 6.1. We say that a representation

$$\bar{\rho} : \text{Gal}(\bar{\mathbb{F}}/\mathbb{F}) \longrightarrow \text{Aut}(V),$$

where V is a 2-dimensional \mathbb{F} -vector space (where \mathbb{F} is a finite field), is *finite at \mathfrak{p}* if the restriction of $\bar{\rho}$ to a decomposition group at \mathfrak{p} is finite, i.e., if there exists a finite flat \mathbb{F} -vector space scheme \mathcal{V} over $\mathcal{O}_{\mathbb{F},\mathfrak{p}}$ such that the resulting representation of $\text{Gal}(\bar{\mathbb{F}}_{\mathfrak{p}}/\mathbb{F}_{\mathfrak{p}})$ coincides with $\bar{\rho}|_{\text{Gal}(\bar{\mathbb{F}}_{\mathfrak{p}}/\mathbb{F}_{\mathfrak{p}})}$.

We prove, in exactly the same way as [19], the following version of Mazur's Principle:

Theorem 6.2 (Mazur's Principle). *Let \mathbb{F} be a totally real number field, and let \mathfrak{p} be a prime of \mathbb{F} dividing the rational prime p . Suppose that*

$$\bar{\rho} : \text{Gal}(\bar{\mathbb{F}}/\mathbb{F}) \longrightarrow \text{GL}_2(\bar{\mathbb{F}}_{\mathfrak{p}})$$

is a continuous irreducible semisimple representation which is attached to a Hilbert cuspidal eigenform $f \in S_{2t,t}(U_0(\mathfrak{p}) \cap U_1(\mathfrak{n}))$, where $\mathfrak{p} \nmid \mathfrak{n}$. Suppose also that

- 1) *If $[\mathbb{F}(\mu_p) : \mathbb{F}] = 2$, then $\bar{\rho}$ is not induced from a character of $\text{Gal}(\bar{\mathbb{F}}/\mathbb{F}(\sqrt{-3}))$.*
- 2) *$e < p - 1$, where e denotes the absolute ramification index of $\mathbb{F}_{\mathfrak{p}}$.*

Then if $\bar{\rho}$ is finite at \mathfrak{p} , then there exists a Hilbert cuspidal eigenform $f' \in S_{2t,t}(U_1(\mathfrak{n}))$ to which $\bar{\rho}$ is attached.

(Note that we will throughout confuse $U_0(\mathfrak{p})$, defined as a subgroup of $\text{GL}_2(\mathcal{O}_{\mathbb{F},\mathfrak{p}})$, with the subgroup of $G(\mathbb{A}^{\infty})$ whose component at \mathfrak{p} is $U_0(\mathfrak{p})$, but which is the fixed maximal open compact subgroup at every place $\mathfrak{q} \neq \mathfrak{p}$, and similarly for other subgroups defined only at a subset of all of the finite places.)

By the first assumption on \mathbb{F} , we may increase the level by replacing $U_1(\mathfrak{n})$ by $H = U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{q}_0)$ for some prime $\mathfrak{q}_0 \nmid \mathfrak{n}\mathfrak{p}$, in such a way that H is sufficiently small in the sense of Theorem 2.1 (and thus Theorem 2.2) and that there are no congruences between forms of level $U_0(\mathfrak{p}) \cap U_1(\mathfrak{n})$ and

forms of level $U_0(\mathfrak{p}) \cap H$ which are new at \mathfrak{q}_0 in the sense that the eigenform does not arise from any level not a multiple of \mathfrak{q}_0 . See [12], §12 for a proof of this fact.

Although the theorem above does not apply when $p = 2$, an analogous result on the existence of auxiliary primes is valid (as long as $\bar{\rho}$ is not induced from a character of $\text{Gal}(\bar{\mathbb{F}}/\mathbb{F}(\sqrt{-1}))$); for this, we can use a generalisation of an argument due to Kevin Buzzard which is proven in [14].

Let $\mathbb{T} = \mathbb{T}_{2t,t,\text{GL}_2}^{S,\mathfrak{p}}(U)$, where $U = U_0(\mathfrak{p}) \cap H$, and where S denotes all primes dividing $n\mathfrak{q}_0$. Let \mathfrak{m} denote the maximal ideal of \mathbb{T} corresponding to $\bar{\rho}$. Then there is a model of $\bar{\rho}$ defined over \mathbb{T}/\mathfrak{m} .

Choose a prime $\mathfrak{q}_1 \nmid n\mathfrak{p}\mathfrak{q}_0$ such that $\bar{\rho}(\text{Frob}_{\mathfrak{q}_1})$ is conjugate to $\bar{\rho}(\sigma)$, where σ denotes complex conjugation. Then if $[\mathbb{F} : \mathbb{Q}]$ is even, the main result of [21] shows that $\bar{\rho}$ is also attached to a Hilbert cuspidal eigenform of level $U_0(\mathfrak{p}\mathfrak{q}_1) \cap H$ of weight $2t$ and which is new at \mathfrak{q}_1 . (The same result is true for $[\mathbb{F} : \mathbb{Q}]$ odd by [18], Theorem 5, but we do not need this here; note, however, that this improves certain results in [13].)

If $[\mathbb{F} : \mathbb{Q}]$ is odd (resp. even), then let B denote the quaternion algebra over \mathbb{F} ramified at $\{\tau_2, \dots, \tau_d\}$ (resp. $\{\mathfrak{q}_1, \tau_2, \dots, \tau_d\}$), and write G for $\text{Res}_{\mathbb{F}/\mathbb{Q}}(B^\times)$. Let $\mathbb{T}' = \mathbb{T}_{2t,t,B}^{S,\mathfrak{p}}(U_0(\mathfrak{p}) \cap H)$ (resp. $\mathbb{T}' = \mathbb{T}_{2t,t,B}^{S \cup \{\mathfrak{q}_1\},\mathfrak{p}}(U_0(\mathfrak{p}) \cap H)$). By [21], there is a maximal ideal \mathfrak{m}' of \mathbb{T}' such that there is a natural isomorphism $\bar{\rho}_{\mathfrak{m}'} \cong \bar{\rho}_{\mathfrak{m}}$.

We use the Jacquet-Langlands correspondence to see that then there is a cuspidal automorphic representation on B of weight $2t$ corresponding to f and with fixed vectors under $U_0(\mathfrak{p}) \cap H$. Let \mathfrak{m}' denote the corresponding maximal ideal of \mathbb{T}' , and let $\mathbb{F} = \mathbb{T}/\mathfrak{m} \cong \mathbb{T}'/\mathfrak{m}'$. Let V be the 2-dimensional \mathbb{F} -vector space corresponding to $\bar{\rho}$. By the results of the previous section, the corresponding maximal ideal of the Hecke algebra is in the support of the p -torsion of the Jacobian, $J_{U_0(\mathfrak{p}),H}[p]$.

Indeed, we see that V is a $\mathbb{F}[\text{Gal}(\bar{\mathbb{F}}/\mathbb{F})]$ -submodule of $J_{U_0(\mathfrak{p}),H}[p]$. As V is finite, it is the generic fibre of a finite flat \mathbb{F} -vector space scheme \mathcal{V} over $\mathcal{O}_{\mathbb{F},\mathfrak{p}}$.

7. Mazur's Principle—proof

After the preliminaries of the previous section, we now give the proof itself.

Lemma 7.1. *The inclusion $\iota : V \hookrightarrow J$ prolongs to an embedding $\iota : \mathcal{V} \hookrightarrow \mathcal{J}$, where \mathcal{J} is the Néron model of the Jacobian $J_{U_0(\mathfrak{p}),H}$ over $\mathcal{O}_{\mathbb{F},\mathfrak{p}}$.*

The proof of this lemma is exactly the same as [19], Lemma 6.2.

Note that if $e = p - 1$, one should still be able to argue as in Edixhoven's paper [9], Theorem 2.8; one needs to know that a Hilbert cuspidal eigenform of weight 2 and level $\mathfrak{p}n$ ($\mathfrak{p} \nmid n$) which is not finite at \mathfrak{p} must be ordinary at \mathfrak{p} . One should then be able to prove results like those of Wiles [22],

Lemma 2.1.5, giving the structure of the local Galois representation at \mathfrak{p} . The remainder of the argument would follow as in [9]. (Thanks to Bas Edixhoven for explaining this argument.)

We recall from [1], [19] or SGA7, the structure of the special fibre $\overline{\mathcal{J}}$ of the Néron model \mathcal{J} of $J_{U_0(\mathfrak{p}),H}$.

Combining the formalism of [19], §2, with Theorem 2.2 (note that $M_{U_0(\mathfrak{p}),H}$ has a regular integral model), we see that $\overline{\mathcal{J}}^0$, the connected component of the special fibre of \mathcal{J} , is an extension of the abelian variety $J_{0,H} \times J_{0,H}$ by a torus $\overline{\mathcal{T}}$ which one can describe explicitly ([19], Proposition 2.1). Let X denote the character group of $\overline{\mathcal{T}}$, and let Φ denote the group of components of $\overline{\mathcal{J}}$ (as we have a regular integral model, an easy calculation shows that $\Phi \cong \mathbb{Z}/s_H\mathbb{Z}$). $\overline{\mathcal{T}}$ lifts to a torus \mathcal{T} over $\mathcal{O}_{F,\mathfrak{p}}$, which embeds in the formal completion of \mathcal{J} along $\overline{\mathcal{J}}$.

Lemma 7.2. $T_{\mathfrak{p}}$ acts on X in the same way as $\text{Frob}_{\mathfrak{p}}$.

Proof. Exactly as in [19], Proposition 3.7, we see that $T_{\mathfrak{p}} + w_{\mathfrak{p}}$ acts trivially on $\overline{\mathcal{T}}$, using Proposition 3.1 above. By the discussion at the end of §2, we know that $w_{\mathfrak{p}}$ acts on Σ_H in the same way as $\text{Frob}_{\mathfrak{p}}$. It follows as in [19], Proposition 3.8, that $w_{\mathfrak{p}}$ acts on X in the same way as $-\text{Frob}_{\mathfrak{p}}$. \square

As in [19], Proposition 5.2(a), we see that $W = J_{U_0(\mathfrak{p}),H}[\mathfrak{m}]$ (semisimple by [2]), regarded as an $\mathbb{F}[\text{Gal}(\overline{F}/F)]$ -module, decomposes as a product of copies of V , the Cartier dual of the group scheme given by $\overline{\rho}$. For this, we note that $\overline{\rho}(\text{Frob}_{\mathfrak{q}})$ satisfies $X^2 - S_{\mathfrak{q}}^{-1}T_{\mathfrak{q}}X + S_{\mathfrak{q}}^{-1}.N_{F/\mathbb{Q}}(\mathfrak{q}) = 0$ for all $\mathfrak{q} \notin S$; as $W \neq 0$ (because \mathfrak{m} is in its support), the argument of [19] extends to this case.

Let us assume that \mathfrak{m} is not in the support of $J_{0,H}[p]$. Let $\overline{\mathcal{V}}$ denote the special fibre of \mathcal{V} . As \mathcal{V} is contained in \mathcal{J} , $\overline{\mathcal{V}}$ is contained in $\overline{\mathcal{J}}$. In our situation, the analogue of [19], Proposition 3.12, is trivial, because our integral model is regular (so, in the notation of [19], $\Lambda = \Lambda^*$, and so clearly $\eta_{\mathfrak{q}}(\Lambda^*) = \eta_{\mathfrak{q}}(\Lambda) \subset X$). It follows that $\overline{\mathcal{V}}$ is contained in $\overline{\mathcal{J}}^0$.

There is an extension

$$0 \longrightarrow \overline{\mathcal{T}} \longrightarrow \overline{\mathcal{J}}^0 \longrightarrow J_{0,H} \times J_{0,H} \longrightarrow 0.$$

By our assumption, $\overline{\mathcal{V}}$ is contained in $\overline{\mathcal{T}}$. The argument of [19], Lemma 6.3, extends to this case, and shows that \mathcal{V} is contained in \mathcal{T} . It follows that

$$V \subset \mathcal{T}[\mathfrak{m}](\overline{F}_{\mathfrak{p}}) = \text{Hom}(X/\mathfrak{m}X, \mu_p).$$

The Frobenius $\text{Frob}_{\mathfrak{p}}$ acts on $X/\mathfrak{m}X$ by $T_{\mathfrak{p}}$, using Lemma 7.2 above; this action is an automorphism, the negative of the automorphism $w_{\mathfrak{p}}$ defined above.

Let $I \subset \text{Gal}(\overline{F}_{\mathfrak{p}}/F_{\mathfrak{p}})$ denote the inertia group at \mathfrak{p} . Then I acts on V by χ , the mod p cyclotomic character.

As V is 2-dimensional, the determinant of the action of I on V is given by χ^2 ; however, we know that the determinant of $\bar{\rho}$ is given by the cyclotomic character χ , by the Čebotarev density theorem. Thus χ must be trivial on inertia. But $e < p - 1$, so $\mu_p \not\subset F_{\mathfrak{p}}$, and this implies that χ is not trivial on inertia. This contradiction shows that the assumption that V is not modular of level \mathfrak{n} is false. Note that if $[F : \mathbb{Q}]$ is even, we need to remove the auxiliary prime \mathfrak{q}_1 added at the end of §6 using Theorem 1.1. \square

8. An application

We end with an application of our main result, together with other known results on level lowering. The argument is exactly the same as Ribet's argument in [19]: we attempt to deduce an analogue of Fermat's Last Theorem over certain totally real number fields from the modularity of semistable elliptic curves. For a study of the (unfortunately rare!) occasions when the hypotheses of this theorem are satisfied in the case of real quadratic fields, see the work with Jayanta Manoharmayum ([15]) and Paul Meekin ([16] and [17]). For simplicity, we will assume that F is a real quadratic field, although similar arguments are applicable for more general totally real fields.

We follow the idea of Frey, Serre and Ribet. Suppose that $\ell \geq 7$ is prime, and that $\alpha^\ell + \beta^\ell = \gamma^\ell$ is a point on the Fermat curve of degree ℓ and with α, β and γ being non-zero elements of a totally real field F . Then we form the Frey curve E :

$$y^2 = x(x - \alpha^\ell)(x + \beta^\ell),$$

which is an elliptic curve over F . We ask whether this curve is modular, in the sense that its L -function agrees with the L -function of some weight 2 Hilbert cusp form over F . Let us make the following two hypotheses:

- 1) Suppose that E is a semistable elliptic curve over F ;
- 2) suppose that $\bar{\rho} = \bar{\rho}_{E,\ell}$ is absolutely irreducible.

As already remarked, we study these hypotheses over real quadratic fields in the papers [16] and [17]. Then we have the following application of the main theorem of the paper:

Theorem 8.1. *Suppose (1) and (2) above. If the Frey curve E is modular, then there is a Hilbert cuspidal eigenform*

$$f \in S_{2t,t}(\Gamma_2^F)$$

such that $\bar{\rho} \cong \bar{\rho}_{f,\lambda}$ for some prime λ of the number field generated by the Hecke eigenvalues of f , where

$$\Gamma_2^F = \bigcap_{\mathfrak{p}|2} U_0(\mathfrak{p}) = U_0\left(\prod_{\mathfrak{p}|2} \mathfrak{p}\right).$$

The proof of this result is again exactly the same as that of Ribet [19]. One considers the mod ℓ representation $\bar{\rho}$ associated to the Frey curve. The discriminant of the Frey curve is $16(\alpha\beta\gamma)^{2\ell}$, and the conductor \mathfrak{n} is the product of the primes dividing $\alpha\beta\gamma$ (as E is semistable); as E is modular, there is a Hilbert cuspidal eigenform f' of weight $(2t, t)$ and level \mathfrak{n} whose Galois representations coincide with those of E . From the theory of Tate curves, we see that if $\mathfrak{p} \nmid 2\ell$ is a prime ideal of \mathcal{O}_F dividing \mathfrak{n} , then $\bar{\rho}$ is unramified at \mathfrak{p} . Further, if $\mathfrak{p} \mid \ell$, then one argues as in Proposition 8.2 of [9] that $\bar{\rho}$ is finite at \mathfrak{p} .

As $[F : \mathbb{Q}]$ is even, we begin by adding an auxiliary prime to the level as in §8 of [13] (which we may do thanks to [21]). (This would not be necessary when $[F : \mathbb{Q}]$ is odd.)

Now we lower the level. As $[F : \mathbb{Q}] = 2$, then we see that if $\ell > 3$, we will certainly have $e < \ell - 1$. Indeed, $\ell \geq 7$, so that also $[F(\mu_\ell) : F] > 2$. Now we can use the result of this paper to remove all primes dividing ℓ from \mathfrak{n} . Next, we may use the results of [12] and [18] to remove the remaining primes not dividing 2 from the level. Finally, we can remove the auxiliary prime which we added by using Fujiwara's version (Theorem 1.1) of Mazur's Principle in the even degree case. The conclusion is that there is some (adelic) Hilbert cuspidal eigenform f of weight $(2t, t)$ and of (semistable) level dividing 2, i.e., on the group given in the statement of the theorem. \square

As a final remark, we should point out that, as in the case of \mathbb{Q} , there is presumably a variant of the main result of this paper in terms of "lowering the weight".

References

- [1] S.Bosch, W.Lütkebohmert, M.Raynaud, Néron Models, Ergebnisse der Mathematik 21, Springer Berlin Heidelberg (1990)
- [2] N.Boston, H.Lenstra, K.Ribet, Quotients of group rings arising from two-dimensional representations, C.R. Acad. Sci. Paris Sér. I Math. 312 (1991) 323–328
- [3] K.Buzzard, On level lowering for mod 2 representations, Math. Res. Lett. 7 (2000) 95–110
- [4] H.Carayol, Sur la mauvaise réduction des courbes de Shimura, Comp. Math. 59 (1986) 151–230
- [5] H.Carayol, Sur les représentations ℓ -adiques associées aux formes modulaires de Hilbert, Ann. Sci. Ec. Norm. Sup. 19 (1986) 409–468
- [6] H.Carayol, Sur les représentations galoisiennes modulo ℓ attachées aux formes modulaires, Duke Math. J. 59 (1989) 785–801
- [7] F.Diamond, R.Taylor, Non-optimal levels of mod ℓ modular representations, Invent. Math. 115 (1994) 435–462
- [8] F.Diamond, R.Taylor, Lifting modular mod ℓ representations, Duke Math. J. 74 (1994) 253–269

- [9] B.Edixhoven, The weight in Serre's conjectures on modular forms, *Invent. Math.* 109 (1992) 563–594
- [10] K.Fujiwara, Level optimisation in the totally real case, preprint
- [11] H.Hida, On p -adic Hecke algebras for GL_2 over totally real fields, *Ann. Math.* 128 (1988) 295–384
- [12] F.Jarvis, Mazur's Principle for totally real fields of odd degree, *Comp. Math.* 116 (1999) 39–79
- [13] F.Jarvis, Level lowering for modular mod ℓ representations over totally real fields, *Math. Ann.* 313 (1999) 141–160
- [14] F.Jarvis, Optimal levels for modular mod 2 representations over totally real fields, preprint (2000)
- [15] F.Jarvis, J.Manoharmayum, On the modularity of elliptic curves over totally real number fields, submitted (2003)
- [16] F.Jarvis, P.Meekin, The Fermat equation over $\mathbb{Q}(\sqrt{2})$, submitted (2003)
- [17] F.Jarvis, P.Meekin, Limitations of the Ribet-Wiles method over real quadratic fields, in preparation
- [18] A.Rajaei, On the levels of Hilbert modular forms, *J. reine angew. Math.* 537 (2001) 33–65
- [19] K.Ribet, On modular representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms, *Invent. Math.* 100 (1990) 431–476
- [20] C.Skinner, A.Wiles, Base change and a problem of Serre, *Duke Math. J.* 107 (2001) 15–25
- [21] R.Taylor, On Galois representations associated to Hilbert modular forms, *Invent. Math.* 98 (1989) 265–280
- [22] A.Wiles, On ordinary λ -adic representations associated to modular forms, *Invent. Math.* 94 (1988) 529–573

DEPARTMENT OF PURE MATHEMATICS
UNIVERSITY OF SHEFFIELD
SHEFFIELD S3 7RH
U.K.

E-mail address: a.f.jarvis@shef.ac.uk
<http://www.shef.ac.uk/~pm1afj/>