Compositio Mathematica **116:** 39–79, 1999.

Mazur's Principle for Totally Real Fields of Odd Degree

FRAZER JARVIS
Mathematical Institute, 24–29 St Giles, Oxford OX1 3LB, U.K.
e-mail: jarvisa@maths.ox.ac.uk

(Received: 29 May 1996; accepted in final form: 15 October 1997)

Abstract. In this paper, we prove an analogue of the result known as Mazur's Principle concerning optimal levels of mod ℓ Galois representations. The paper is divided into two parts. We begin with the study (following Katz–Mazur) of the integral model for certain Shimura curves and the structure of the special fibre. It is this study which allows us to generalise, in the second part of this paper, Mazur's result to totally real fields of odd degree.

Mathematics Subject Classifications (1991): 11F33, 11F41, 11F80, 11G18, 14G35.

Key words: modular forms, Galois representations, totally real fields.

1. Introduction

In [24], Serre makes a series of conjectures concerning mod ℓ Galois representations, that is, representations $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$. In particular, Serre defines a notion of 'modularity'; $\overline{\rho}$ is *modular* if it is the reduction of a representation associated to a modular form (see [8]), and he predicts a criterion to determine whether a given $\overline{\rho}$ is modular. In general, a modular representation will be modular in many ways, so that there will be many modular forms such that the reduction of the associated Galois representation is isomorphic to $\overline{\rho}$. Amongst these, Serre predicts an 'optimal' weight and level for such a form. That these are correct is now known, at least if ℓ is odd (see [10]). Serre also speculates on the possibility of a 'mod ℓ version of the Langlands philosophy'.

Carayol [4] and Taylor [25] prove that to any Hilbert cusp form over a totally real field F, which is an eigenform for all of the Hecke operators in a certain Hecke algebra, one may attach representations of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$ in a similar way to the classical case. We may thus begin to think about mod ℓ representations $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ and begin the study of modularity in this new context.

In this paper, we prove the following result on lowering the level of a mod ℓ representation $\overline{\rho}$. This result is analogous to Mazur's Principle (see [23]), one of the main ingredients in Ribet's work (when $\mathrm{F} = \mathbf{Q}$) leading to the proof that Serre's optimal level is the correct one.

INTERPRINT: J.N.B. PIPS Nr.: 152827 MATHKAP
comp4086.tex; 7/08/1995; 8:16; v.7; p.1

Throughout, our notation for Hilbert modular forms will follow that of Hida [15].

THEOREM (Mazur's Principle). *Assume* $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$, *a continuous irreducible semisimple representation, is attached to a Hilbert cuspidal eigenform* $f \in S_{k,w}(U_0(\mathfrak{p})\cap U_1(\mathfrak{n}))$, *where* $\mathfrak{p} \nmid \mathfrak{n}\ell$, *and* $k \geqslant 2t$. *Suppose* $[\mathrm{F}(\mu_\ell)\colon \mathrm{F}] \geqslant 4$. *If* $\mathrm{F}/\mathbf{Q}$ *has even degree, suppose also that there exists some finite place* $\mathfrak{q}_0 \neq \mathfrak{p}$ *of* F *at which the automorphic representation corresponding to* f *is special or supercuspidal. Then if* $\overline{\rho}$ *is irreducible, and unramified at* $\mathfrak{p}$, *and* $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$, *there exists a Hilbert cuspidal eigenform* $f' \in S_{k,w}(U_1(\mathfrak{n}))$ *to which* $\overline{\rho}$ *is attached.*

As F is totally real, the degree of the extension $\mathrm{F}(\mu_\ell)/\mathrm{F}$ is always even (F is contained in the maximal totally real subfield $\mathrm{F}(\mu_\ell)^+$ of $\mathrm{F}(\mu_\ell)$ and the degree of $\mathrm{F}(\mu_\ell)/\mathrm{F}(\mu_\ell)^+$ is 2). Thus the condition on F and ℓ holds unless $[\mathrm{F}(\mu_\ell)\colon \mathrm{F}] = 2$; in this case, one can use the enhancement of Diamond and Taylor (see for instance [7], Lemma 4.11).

In another paper, we derive analogues of results of Carayol on modular mod ℓ representations, and we intend to investigate analogues of Ribet's theorem in the future.

The proof of Mazur's Principle is geometric in nature, and involves a deep study of certain modular curves (see [9] and [17]). In the first half of this paper, we make an analogous study of the corresponding Shimura curves (made slightly harder as these Shimura curves have no natural interpretation as the moduli space of abelian varieties), and prove our main result in the second half of the paper. Whilst the results in the first half of the paper are valid for any totally real field, the results concerning Hilbert modular forms in the second half of the paper are only valid when the hypotheses of the above theorem hold (but certainly include all totally real fields of odd degree). We hope that the study carried out in the first half of this paper will be more generally useful in the study of mod ℓ Galois representations over totally real fields.

2. Notation

Our notation will follow that of [3].

Let F be a totally real field of degree d over $\mathbf{Q}$. Denote by $\tau_1, \ldots, \tau_d$ the infinite places of F, and let B be a quaternion algebra over F split at exactly one infinite place, τ_1, say. Let $\mathfrak{p}$ be a finite place of F at which B splits; let κ denote the residue field, with cardinality q and characteristic p. Write $\mathcal{O}_\mathfrak{p}$ for the ring of integers of $\mathrm{F}_\mathfrak{p}$, and we denote also by $\mathfrak{p}$ a uniformiser. As usual, we write $\mathcal{O}_{(\mathfrak{p})}$ for $\mathrm{F}\cap\mathcal{O}_\mathfrak{p}$ and $\widehat{\mathcal{O}}_\mathfrak{p}^{\mathrm{nr}}$ for the completion of the ring of integers of $\mathrm{F}_\mathfrak{p}^{\mathrm{nr}}$. We fix an isomorphism between $(B\otimes_\mathrm{F}\mathrm{F}_v)^\times$ and $\mathrm{GL}_2(\mathrm{F}_v)$ at places v at which B is split.

Define $G = \mathrm{Res}_{\mathrm{F}/\mathbf{Q}}(B^\times)$. Then, if K denotes a compact open subgroup of $G(\mathbf{A}^\infty)$, where $\mathbf{A}^\infty$ denotes the *finite* adeles (of $\mathbf{Q}$), we define the associated *Shimura curve* by $M_K(\mathbf{C}) = G(\mathbf{Q})\backslash(G(\mathbf{A}^\infty)/K \times (\mathbf{C}-\mathbf{R}))$. Then, by work of

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

Shimura, $M_K(\mathbf{C})$ has a canonical model, denoted M_K, over F (see [3]). Our aim in the first half of this paper will be to demonstrate the existence of a scheme $\mathbf{M}_K$, defined over $\mathcal{O}_{(\mathfrak{p})}$, such that there is an isomorphism of F-schemes $\mathbf{M}_K \otimes_{\mathcal{O}_{(\mathfrak{p})}} \mathrm{F} \cong M_K$, whenever $K = K_{\mathfrak{p}} K^{\mathfrak{p}}$ with $K^{\mathfrak{p}}$ sufficiently small, and $K_{\mathfrak{p}}$ is one of the groups

$$U_1(\mathfrak{p}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}}) \middle| a - 1 \in \mathfrak{p}, c \in \mathfrak{p} \right\},$$

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

We will also be able to study the special fibres of these models. Write also

$$U_n = \{\alpha \in \mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}}) | \alpha \equiv I_2 \pmod{\mathfrak{p}^n}\}.$$

Following Carayol, we write H for $K^{\mathfrak{p}}$, and write $M_{0,H}$ (resp. $M_{n,H}$) to abbreviate $M_{\mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}}) \times H}$ (resp. $M_{U_n \times H}$). In the same way, we will write

$$M_{U_0(\mathfrak{p}),H} = M_{U_0(\mathfrak{p}) \times H}, \qquad M_{U_1(\mathfrak{p}),H} = M_{U_1(\mathfrak{p}) \times H}.$$

3. The Results of Carayol

For proofs or more details for all of the facts in this section, we refer the reader to [3].

In [3], Carayol gives a proof of the following result, first proven by Morita.

THEOREM 3.1. *Suppose H is sufficiently small. There exists an integral model $\mathbf{M}_{0,H}$, defined over $\mathcal{O}_{(\mathfrak{p})}$, for $M_{0,H}$. This model has good reduction, in that it is proper and smooth.*

However, in order to study the bad reduction of Shimura curves when $K_{\mathfrak{p}}$ is more complicated, Carayol introduces some more machinery.

If the model $\mathbf{M}_{0,H}$ were to be the solution to a moduli problem of Abelian varieties, then we would have a universal Abelian variety on which we could impose some extra conditions, in order to study moduli problems with nontrivial level structures at $\mathfrak{p}$. However, $\mathbf{M}_{0,H}$ is not the solution to any such moduli problem. But Carayol is able to define some substitute objects – for each H, there exists an integer n, such that one can define finite flat group schemes $\mathbf{E}_{1,H}, \ldots, \mathbf{E}_{n,H}$ on $\mathbf{M}_{0,H}$, such that $\mathbf{E}_{i,H}$ has rank q^{2i}, and has an action of $\mathcal{O}_{\mathfrak{p}}$. These group schemes play the part of the $\mathfrak{p}^n$-torsion points on a universal abelian variety, and we can use these to give the definitions of level structures at $\mathfrak{p}$.

For each subgroup H, one can only define finitely many of these finite group schemes. However, as H gets smaller and smaller, increasingly many of them are defined, so that, on the projective limit $\mathbf{M}_0$ of the system $\{\mathbf{M}_{0,H}\}$, we recover a full *divisible $\mathcal{O}_{\mathfrak{p}}$-module* $\mathbf{E}_\infty$ (that is, a one-dimensional p-divisible group with an

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

action of $\mathcal{O}_\mathfrak{p}$). For a summary of Drinfeld's theory of divisible $\mathcal{O}_\mathfrak{p}$-modules, see the appendix to [3].

Furthermore, at every geometric point x of the special fibre $\mathbf{M}_{0,H} \otimes \kappa$, one can define a *local* divisible $\mathcal{O}_\mathfrak{p}$-module, although these do not glue together in a compatible manner. One considers the étale covering $\mathbf{M}_0 \longrightarrow \mathbf{M}_{0,H}$, and chooses a lift y of x. The map then gives an isomorphism between the local ring at y and the local ring at x. One then considers the pull-back of the divisible $\mathcal{O}_\mathfrak{p}$-module $\mathbf{E}_\infty$ over $\mathbf{M}_0$ via the morphism $y\colon \operatorname{spec} \overline{\kappa} \longrightarrow \mathbf{M}_0$. Any two choices of y give rise to isomorphic divisible 'height 2' $\mathcal{O}_\mathfrak{p}$-modules over $\overline{\kappa}$, and we write $\mathbf{E}_\infty|_x$ for the result.

Drinfeld [13] has given a classification of divisible height 2 $\mathcal{O}_\mathfrak{p}$-modules over $\overline{\kappa}$, and has shown that there exists, up to isomorphism, a unique formal $\mathcal{O}_\mathfrak{p}$-module of each height h, Σ_h say. It follows that there are, up to isomorphism, just two possibilities for $\mathbf{E}_\infty|_x$:

(1) $\mathbf{E}_\infty|_x \cong \Sigma_1 \times (\mathrm{F}_\mathfrak{p}/\mathcal{O}_\mathfrak{p})$,
(2) $\mathbf{E}_\infty|_x \cong \Sigma_2$.

We call x *ordinary* if the first case occurs, and *supersingular* if the second case occurs. Carayol [3] proves that the set

$$\{x \text{ a geometric point of } \mathbf{M}_{0,H} \otimes \kappa |\ \mathbf{E}_\infty|_x \text{ is supersingular}\}$$

is a finite nonempty set.

Carayol constructs a scheme $\mathbf{M}_{n,H}$ as 'the moduli space of full level $\mathfrak{p}^n$ structures' (adapting a definition of Drinfeld [13] of such structures) on $\mathbf{M}_{0,H}$, and proves that it is an integral model for the Shimura curve $M_{n,H}$. He is then able to analyse this integral model, and it is this study which plays the key role in his proof [4] that one may attach Galois representations to Hilbert modular forms of weight $k \geqslant 2t$ whenever F is an extension of $\mathbf{Q}$ of odd degree.

For the applications we have in mind to the theory of modulo ℓ Galois representations attached to Hilbert modular forms, it will only be necessary for us to extend these ideas to give integral models when $K_\mathfrak{p}$ is either $U_0(\mathfrak{p})$ or $U_1(\mathfrak{p})$, but it is likely that the methods of [17] would prove the existence of integral models when $K_\mathfrak{p}$ is any subgroup of $\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})$.

4. Deformation Theory I

We begin the study of certain deformation problems in this section; this study will be important later to analyse the local rings of the integral models that we shall construct.

DEFINITION 4.1. Let $\mathcal{C}$ denote the category of complete noetherian local $\widehat{\mathcal{O}}_\mathfrak{p}^{\mathrm{nr}}$-algebras with residue field $\overline{\kappa}$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

Throughout this section, E will denote a divisible $\mathcal{O}_{\mathfrak{p}}$-module of height 2 over a ring $R \in \mathcal{C}$ whose maximal ideal we denote by $\mathfrak{m}_R$. Its connected part, E^0, is thus equipped with a 'formal multiplication law', $f\colon \mathcal{O}_{\mathfrak{p}} \longrightarrow \operatorname{End} E^0$. Identifying E^0 with spf $R[[X]]$, formal multiplication by $a \in \mathcal{O}_{\mathfrak{p}}$ is given by a power series $f_a(X) \in R[[X]]$.

DEFINITION 4.2. Let E be supersingular (i.e., $E = E^0$). A $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-*map* on E is a homomorphism of $\mathcal{O}_{\mathfrak{p}}$-modules

$$\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}}) \longrightarrow \operatorname{Hom}(\operatorname{spf} R, E) = \mathfrak{m}_R,$$

(where $\mathfrak{m}_R$ inherits the structure of an $\mathcal{O}_{\mathfrak{p}}$-module from E) such that $f_{\mathfrak{p}}(X)$ is divisible by $\prod_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})}(X - \phi(\alpha))$ in $R[[X]]$.

DEFINITION 4.3. Let E be supersingular. A $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-*structure* on E is an equivalence class of $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-maps, where

$$\phi \sim \phi' \iff \prod_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})} (X - \phi(\alpha)) = \prod_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})} (X - \phi'(\alpha)).$$

When $R = \overline{\kappa}$, there is a unique $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-map (and consequently a unique $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-structure) as $\Sigma_2(\overline{\kappa}) = \{0\}$.

DEFINITION 4.4. Let (E, f) be a divisible $\mathcal{O}_{\mathfrak{p}}$-module of height 2 over $\overline{\kappa}$. Let $R \in \mathcal{C}$. Then a *deformation* of (E, f) to R is a divisible $\mathcal{O}_{\mathfrak{p}}$-module $(\widetilde{E}, \widetilde{f})$ over R whose reduction modulo $\mathfrak{m}_R$ is (E, f). If also E is supersingular, we say that a *deformation with* $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-*map* is a deformation $(\widetilde{E}, \widetilde{f})$, together with a $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-map $\widetilde{\phi}$ on $\widetilde{E}$.

The following result is implicit in the proof of the lemma after Proposition 4.3 of [13].

THEOREM 4.5 (Drinfeld). *Let (E, f) be a divisible height 2 $\mathcal{O}_{\mathfrak{p}}$-module over $\overline{\kappa}$.*

(1) *The functor which to each $R \in \mathcal{C}$ associates the set of isomorphism classes of deformations of (E, f) to R is represented by a ring D_0^E, isomorphic to $\widehat{\mathcal{O}}_{\mathfrak{p}}^{\mathrm{nr}}[[t_1]]$.*
(2) *If also E is supersingular, then the functor which to $R \in \mathcal{C}$ associates the set of deformations of (E, f) with $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-map to R is represented by a ring L_1^E, isomorphic to $D_0^E[[y_1]]/(f_{\mathfrak{p}}(y_1)/y_1)$. Further, the ring L_1^E is regular (where $\{y_1, t_1\}$ form a regular sequence of parameters), and the morphism $D_0^E \longrightarrow L_1^E$ is finite and flat.*

We are also interested in the deformation problem for $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-structures. Note that the deformation ring can almost certainly be directly computed using Cartier-Dieudonné theory, as in [21], ch. 2, but we will instead adapt the method of [17].

5. Deformation Theory II

In this section, we gather together the results that we will later use to demonstrate the existence of integral models of Shimura curves with $\Gamma_0(\mathfrak{p})$-structure.

Throughout this section, let E be a supersingular (height 2) divisible $\mathcal{O}_{\mathfrak{p}}$-module over $R \in \mathcal{C}$.

DEFINITION 5.1. A $\Delta(\mathfrak{p})$-*structure* on E is a pair of $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-maps (ϕ, ϕ') such that

$$\prod_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})}(X-\phi(\alpha)) = \prod_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})}(X-\phi'(\alpha)).$$

Thus such a structure consists of a pair giving rise to the same $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-structure.

Define the notion of a deformation with $\Delta(\mathfrak{p})$-structure in the analogous way to Definition 4.4.

THEOREM 5.2. *Let E be a supersingular $\mathcal{O}_{\mathfrak{p}}$-module over $\overline{\kappa}$. The functor which to each D_0^E-algebra $R \in \mathcal{C}$ associates the set of deformations of E with $\Delta(\mathfrak{p})$-structure to R is represented by a ring M_1^E.*

Proof. We imitate [17], (6.3.2). Let $R_1^E = L_1^E \otimes_{D_0^E} L_1^E = L_1^E[[y_2]]/(f_{\mathfrak{p}}(y_2)/y_2)$ be the ring representing pairs of $(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})$-maps.

Let $\Sigma_j^{(q)}$ denote the jth symmetric polynomial on q variables. Then to say that (ϕ, ϕ') form a $\Delta(\mathfrak{p})$-structure on E is equivalent to saying that

$$\Sigma_j^{(q)}(\{\phi(\alpha)\}_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})}) = \Sigma_j^{(q)}(\{\phi'(\alpha)\}_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})})$$

for all $j = 1, 2, \ldots, q = N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p})$. (This equality expresses the condition that the coefficients of X^{q-j} of the relationship defining $\Delta(\mathfrak{p})$-structures are equal.) Then define

$$I_M = \left(\left\{\Sigma_j^{(q)}(\{f_\alpha(y_1)\}_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})}) - \Sigma_j^{(q)}(\{f_\alpha(y_2)\}_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})})\right\}_{j=1,\ldots,q}\right),$$

an ideal of R_1^E.

Let $M_1^E = R_1^E/I_M$. It is easy to see that this ring represents the functor of $\Delta(\mathfrak{p})$-structures.

We now concentrate on constructing integral models for Shimura curves.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

6. Preliminaries

Let H be sufficiently small that $\mathbf{E}_{1,H}$ is defined on $\mathbf{M}_{0,H}$. We will usually just write $\mathbf{E}_1$ for the group scheme $\mathbf{E}_{1,H}$. If S is a scheme over $\mathbf{M}_{0,H}$, then we can consider the pull-back $\mathbf{E}_1|_S$ to S. It is a finite locally free (and hence flat) group scheme over S of rank q^2 and has an action of $\mathcal{O}_\mathfrak{p}$.

DEFINITION 6.1. If a finite locally free scheme has an action of $\mathcal{O}_\mathfrak{p}$, we will refer to it as an *$\mathcal{O}_\mathfrak{p}$-scheme*. A finite locally free group scheme with an $\mathcal{O}_\mathfrak{p}$-action will be called an *$\mathcal{O}_\mathfrak{p}$-group scheme*.

DEFINITION 6.2. By a *set of sections* of some finite locally free scheme Z/S, we will mean an unordered set $\{P_1, \ldots, P_r\}$ (for some r) consisting of not necessarily distinct sections of Z over S.

Notation 6.3. For P a section of $\mathbf{E}_1$ over S, we denote by $[P]$ the subscheme of $\mathbf{E}_1|_S$ which it defines (its image), and $\mathcal{I}_P$ the ideal sheaf defining this subscheme. If $\mathcal{K}$ is a set of sections of $\mathbf{E}_1$ over S, write

$$\mathcal{I}_\mathcal{K} = \prod_{P\in\mathcal{K}} \mathcal{I}_P,$$

and $\sum_{P\in\mathcal{K}}[P]$ for the subscheme of $\mathbf{E}_1|_S$ defined by the ideal sheaf $\mathcal{I}_\mathcal{K}$.

LEMMA 6.4. *Let $\mathcal{K} = \{P_1, \ldots, P_k\}$ be a set of sections for $\mathbf{E}_1$ over S. Then $\sum_{i=1}^k [P_i]$ is a finite subscheme of finite presentation of $\mathbf{E}_1|_S$.*

Proof. Write Z for $\sum_{i=1}^k [P_i]$. The question as to whether Z is a finite subscheme of finite presentation is a local one.

But $\mathbf{E}_1|_S$ is locally free, so S can be covered with open affines over which $\mathbf{E}_1|_S$ is free. Thus we may assume that $S = \operatorname{spec} R$, and $\mathcal{O}_{\mathbf{E}_1|_S} \cong R^{q^2}$.

Let $I_\mathcal{K}$ be the ideal defining Z inside $\mathbf{E}_1|_R$. Then

$$0 \longrightarrow I_\mathcal{K} \longrightarrow R^{q^2} \longrightarrow \mathcal{O}_Z|_R \longrightarrow 0$$

so that $\mathcal{O}_Z|_R$ is clearly finite.

Further, the ideal I_{P_i} defining the subscheme $[P_i]$ is clearly finitely generated, as we have an exact sequence

$$0 \longrightarrow I_{P_i} \longrightarrow R^{q^2} \longrightarrow \mathcal{O}_{[P_i]}|_R \longrightarrow 0,$$

and $\mathcal{O}_{[P_i]}|_R \cong R$ (see [19], 2.6). Thus $I_\mathcal{K}$ is finitely generated, so that, in the first sequence above, $\mathcal{O}_Z|_R$ is of finite presentation ([19], p. 14).

DEFINITION 6.5. We say that a set of sections $\mathcal{K} = \{P_1, \ldots, P_k\}$ is a *subset* (resp. *subgroup of sections*) of $\mathbf{E}_1|_S$ if $\sum_{P\in\mathcal{K}}[P]$ is a subscheme (resp. subgroup scheme) of $\mathbf{E}_1|_S$ which is locally free of rank k.

Remark. Presumably a subset (in the usual sense) of a subset (in this sense) is again a subset (in this sense). We do not know a proof, but we do not need it in this paper.

Note that if P is a section of $\mathbf{E}_1$ over S, then the set $\{P\}$ is a subset of $\mathbf{E}_1|_S$ because $\mathcal{O}_{[P]}|_S \cong \mathcal{O}_S$.

If P is a section of $\mathbf{E}_1|_S$, write $\langle P\rangle$ for the set of sections $\{\lambda P | \lambda \in (\mathcal{O}_\mathfrak{p}/\mathfrak{p})\}$.

We now prove some representability results which we shall use later.

LEMMA 6.6. *Let S be an $\mathbf{M}_{0,H}$-scheme, and let $\mathcal{K}$ be a set of sections of $\mathbf{E}_1$ over S. Then there exists a closed subscheme T of S such that for every morphism $S' \longrightarrow S$, the restriction $\mathcal{K}'$ of $\mathcal{K}$ to S' is a subset of $\mathbf{E}_1|_{S'}$ if and only if $S' \longrightarrow S$ factors through T.*

Proof. The problem in the question is Zariski local, so we may assume S affine, $S = \operatorname{spec} R$. Write Z for the subscheme of $\mathbf{E}_1|_S$ defined by the ideal sheaf $\mathcal{I}_\mathcal{K}$. As Z is of finite presentation, it suffices to consider R noetherian (this is a standard reduction, as in [EGA IV], 8.9.1 and 11.2.6.1).

We may think of Z as $\operatorname{spec} \mathcal{F}$, where $\mathcal{F}$ is a coherent sheaf of algebras on S. Write $N = |\mathcal{K}|$. Then the condition on an S-scheme S' that $\mathcal{K}'$ be a subset of $\mathbf{E}_1|_{S'}$ is that $Z_{S'} = Z \times_S S'$ be a finite locally free S'-scheme of rank N.

For any field valued point $\operatorname{spec} k \longrightarrow S$ of S, the fibre $\mathcal{F} \otimes k$ is a k-vector space such that $\dim_k(\mathcal{F} \otimes k) =$ rank of Z_k, as $\mathcal{F} \otimes k$ is the affine ring of Z_k.

For all such points, this dimension is at most N. This is clear if p is invertible in k, as then Z_k is étale; if k has characteristic p, it suffices to verify the condition at closed points, and one easily verifies the claim at both ordinary and supersingular points.

We now apply Mumford's flattening stratification ([17], 6.4.3) which states:

Given a noetherian scheme S, a coherent sheaf $\mathcal{F}$ on S and an integer N such that for all points $s \in S$, $\dim_{k(s)}(\mathcal{F} \otimes k(s)) \leqslant N$, the condition on S-schemes $S' \longrightarrow S$ that $\mathcal{F}_{S'}$ be locally free of rank N on S' is represented by a closed subscheme T of S.

The lemma is now merely a special case of this result.

DEFINITION 6.7. Let $\mathcal{K}$ and $\mathcal{L}$ be two subsets of $\mathbf{E}_1|_S$. Write $\mathcal{K} \subset \mathcal{L}$ to indicate that $\mathcal{I}_\mathcal{L} \subset \mathcal{I}_\mathcal{K}$.

LEMMA 6.8. *Let S be an $\mathbf{M}_{0,H}$-scheme, and let $\mathcal{K}$ and $\mathcal{L}$ be two subsets of $\mathbf{E}_1|_S$. Then there exists a closed subscheme T of S such that for every morphism $S' \longrightarrow S$, the restrictions $\mathcal{K}'$ and $\mathcal{L}'$ of $\mathcal{K}$ and $\mathcal{L}$ to S' satisfy $\mathcal{K}' \subset \mathcal{L}'$ if and only if the morphism $S' \longrightarrow S$ factors through T.*

Proof. (cf. [17], 6.7.3) For the same reasons as above, one may reduce to the case where S is affine and noetherian. Then if $\mathcal{F}$ denotes the coherent sheaf of

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

algebras defining $\mathbf{E}_1|_S$, the condition that $\mathcal{I}_{\mathcal{L}'} \subset \mathcal{I}_{\mathcal{K}'}$ is satisfied on the closed subscheme of S over which the composite morphism $\mathcal{I}_{\mathcal{L}} \hookrightarrow \mathcal{F} \twoheadrightarrow \mathcal{F}/\mathcal{I}_{\mathcal{K}}$ between locally free sheaves vanishes.

We note that it then follows that the locus where $\mathcal{K}' = \mathcal{L}'$ is also closed.

LEMMA 6.9. *Let a set of sections $\mathcal{K}$ be a subset of $\mathbf{E}_1|_S$. There exists a closed subscheme S_{sub} of S such that if $S' \longrightarrow S$, we have $\mathcal{K}'$ is a subgroup of sections of $\mathbf{E}_1|_{S'}$ if and only if $S' \longrightarrow S$ factors through S_{sub}.*

Proof. This is identical to [17], 1.3.7. Let

$e\colon S \longrightarrow \mathbf{E}_1|_S$ denote the identity section,

$i\colon \mathbf{E}_1|_S \longrightarrow \mathbf{E}_1|_S$ denote inversion,

$m\colon \mathbf{E}_1|_S \times_S \mathbf{E}_1|_S \longrightarrow \mathbf{E}_1|_S$ denote multiplication.

The closed subscheme S_{sub} of S is the locus over which

(1) $\{e\} \subset \mathcal{K}$
(2) $\mathcal{K}$ is closed under inversion, i.e., that $i(\mathcal{K}) = \mathcal{K}$.
(3) $\mathcal{K}$ is closed under multiplication. Let V be the subscheme of $\mathbf{E}_1|_S$ which represents the condition that $\{P\} \subset \mathcal{K}$. Let $W = V \times_S V$, which represents the functor 'ordered pairs of sections in $\mathcal{K}$'. If (P, Q) is the universal pair of sections of $\mathcal{K}$, we require that $\{m(P, Q)\} \subset \mathcal{K}_W$.

(1) and (2) are closed conditions by the previous lemma, and are succesively defined locally on S by finitely many equations. (3) is a closed condition on W, defined locally on W by finitely many equations. But W is locally free over S, so that (3) is also defined locally on S by a finite number of equations (the co-ordinate functions of the equations over W).

LEMMA 6.10. *Let a set of sections $\mathcal{K}$ be a subset of $\mathbf{E}_1|_S$. Then there exists a closed subscheme S_{act} of S such that if $S' \longrightarrow S$, we have an action of $\mathcal{O}_{\mathfrak{p}}$ on $\mathcal{K}'$ if and only if $S' \longrightarrow S$ factors through S_{act}.*

Proof. We require that $\{\lambda P'\} \subset \mathcal{K}'$ for all $P \in \mathcal{K}$ and all $\lambda \in (\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})$. This involves finitely many conditions, each closed.

7. The $U_1(\mathfrak{p})$-Problem

We now construct an integral model for $M_{U_1(\mathfrak{p}),H}$.

DEFINITION 7.1. A $U_1(\mathfrak{p})$-*structure* on S, an $\mathbf{M}_{0,H}$-scheme, is a morphism $\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}}) \longrightarrow \mathrm{Hom}(S, \mathbf{E}_1)$ such that $\sum_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_{\mathfrak{p}})}[\phi(\alpha)]$ is a finite locally free $\mathcal{O}_{\mathfrak{p}}$-subgroup scheme of rank q of $\mathbf{E}_1|_S$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

In the language of Section 6, this condition is equivalent to the following: a $U_1(\mathfrak{p})$-structure is a section P of $\mathbf{E}_1$ over S such that the set of sections $\langle P\rangle$ forms an $\mathcal{O}_\mathfrak{p}$-subgroup of sections of $\mathbf{E}_1|_S$.

DEFINITION 7.2. We define the functor $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$ on the category of $\mathbf{M}_{0,H}$-schemes by defining $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}(S)$ to be the set of $U_1(\mathfrak{p})$-structures on S.

Define a functor $\mathcal{M}_{U_1(\mathfrak{p}),H}$ on $M_{0,H}$-schemes by setting $\mathcal{M}_{U_1(\mathfrak{p}),H}(S)$ to be the set of sections P of $E_1|_S$ which is nowhere the zero section.

LEMMA 7.3. *The two functors* $\underline{\boldsymbol{\mathcal{M}}}_{U_1(\mathfrak{p}),H}$ *and* $\mathcal{M}_{U_1(\mathfrak{p}),H}$ *defined above coincide on* $M_{0,H}$*-schemes.*

Proof. To give an $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$-structure on an $M_{0,H}$-scheme S is to give a section P such that $\sum_{\lambda\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})}[\lambda P]$ is a finite flat subgroup scheme of $\mathbf{E}_1|_S$ of rank q.

But if S is a scheme over $M_{0,H}$, $\mathbf{E}_1|_S = E_1|_S$ which is étale over S. Thus $\mathbf{E}_1|_S$ is, locally in the étale topology, non-canonically isomorphic to $(\mathcal{O}_\mathfrak{p}/\mathfrak{p})^2$, and so the finite flat $\mathcal{O}_\mathfrak{p}$-subgroup schemes of rank q are (locally in the étale topology) isomorphic to $(\mathcal{O}_\mathfrak{p}/\mathfrak{p})$, so that P must be a nowhere zero section killed by multiplication by $\mathfrak{p}$.

LEMMA 7.4. *The functor* $\mathcal{M}_{U_1(\mathfrak{p}),H}$ *is representable by the F-scheme* $M_{U_1(\mathfrak{p}),H}$.

Proof. Define the moduli problem $\mathcal{M}_{1,H}$ on $M_{0,H}$-schemes as in [3], 7.1, so that $\mathcal{M}_{1,H}(S)$ consists of pairs (P,Q) of sections of $E_1|_S$ over S which trivialise $E_1|_S$. It is represented by $M_{1,H}$. That is, there exist a universal pair of sections (P_u,Q_u) of sections of $E_1|_{M_{1,H}}$ over $M_{1,H}$ such that if $(P,Q)\in\mathcal{M}_{1,H}(S)$, there exists a unique morphism of $M_{0,H}$-schemes $S\xrightarrow{\alpha}M_{1,H}$ such that (P,Q) is the image of (P_u,Q_u) under the induced map $\alpha^*\colon\mathcal{M}_{1,H}(M_{1,H})\longrightarrow\mathcal{M}_{1,H}(S)$. In particular, to give an element of $\mathcal{M}_{1,H}(S)$ uniquely specifies an element of $\mathrm{Hom}_{(Sch/M_{0,H})}(S,M_{1,H})$.

If S is an object of $(Sch/M_{0,H})$, then $\mathcal{M}_{1,H}(S)$ has an action of $\mathrm{GL}_2(\mathcal{O}_\mathfrak{p}/\mathfrak{p})$ which is given by

$$(P,Q)\mapsto(P,Q)\begin{pmatrix}a&b\\c&d\end{pmatrix}=(aP+cQ,bP+dQ).$$

Further, if $S\longrightarrow S'$ is a morphism in $(Sch/M_{0,H})$, then the induced map $\mathcal{M}_{1,H}(S')\longrightarrow\mathcal{M}_{1,H}(S)$ is equivariant for this action.

The equivalence classes under the action of the subgroup

$$\widetilde{U}_1(\mathfrak{p})=\left\{\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{GL}_2(\mathcal{O}_\mathfrak{p}/\mathfrak{p})\,\middle|\,a-1\in\mathfrak{p},c\in\mathfrak{p}\right\}$$

are clearly non-zero sections P which are killed by multiplication by $\mathfrak{p}$. Thus we get a map of moduli problems $\mathcal{M}_{1,H}\longrightarrow\mathcal{M}_{U_1(\mathfrak{p}),H}$. Locally for the étale topology, any $U_1(\mathfrak{p})$-structure P may be completed to a pair $(P,Q) \in \mathcal{M}_{1,H}(S)$. To give a trivialisation (P,Q) of $E_1|_S$ for an $M_{0,H}$-scheme S is to give a morphism $S \longrightarrow M_{1,H}$ of $M_{0,H}$-schemes. It follows that to give a $U_1(\mathfrak{p})$-structure on S is to give a section $S \longrightarrow M_{1,H}/\widetilde{U}_1(\mathfrak{p})$. But $M_{1,H}/\widetilde{U}_1(\mathfrak{p}) = M_{U_1(\mathfrak{p}),H}$. Thus, it suffices to give a morphism $S \longrightarrow M_{U_1(\mathfrak{p}),H}$ of $M_{0,H}$-schemes and so it follows that $M_{U_1(\mathfrak{p}),H}$ represents the $\mathcal{M}_{U_1(\mathfrak{p}),H}$-moduli problem.

Thus in showing that $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$ is representable, we are constructing an $\mathcal{O}_{(\mathfrak{p})}$-scheme $\mathbf{M}_{U_1(\mathfrak{p}),H}$ which is an integral model for $M_{U_1(\mathfrak{p}),H}$ in that $\mathbf{M}_{U_1(\mathfrak{p}),H} \otimes_{\mathcal{O}_{(\mathfrak{p})}} \mathrm{F} = M_{U_1(\mathfrak{p}),H}$.

LEMMA 7.5. *$\underline{\boldsymbol{\mathcal{M}}}_{U_1(\mathfrak{p}),H}$ is representable.*

Proof. For an $\mathbf{M}_{0,H}$-scheme S, let $\mathcal{H}om(S)$ be defined as the set of all $\mathcal{O}_\mathfrak{p}$-homomorphisms $\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}) \longrightarrow \mathrm{Hom}(S,\mathbf{E}_1)$. This is clearly represented by $\mathbf{E}_1$. $\phi \in \mathcal{H}om(S)$ is a $U_1(\mathfrak{p})$-structure if the subscheme $\sum_{\alpha\in(\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p})}[\phi(\alpha)]$ of $\mathbf{E}_1|_S$ is a finite subscheme of finite presentation which is a locally free rank q group scheme with an action of $\mathcal{O}_\mathfrak{p}$. $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$ is a closed subfunctor of $\mathcal{H}om$ by Lemmas 6.6, 6.9 and 6.10, and is thus representable by a closed subscheme of $\mathbf{E}_1$.

Let $\mathbf{M}_{U_1(\mathfrak{p}),H}$ represent $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$.

THEOREM 7.6. *The scheme $\mathbf{M}_{U_1(\mathfrak{p}),H}$ is a regular scheme of dimension* 2. *The projection morphism $\mathbf{M}_{U_1(\mathfrak{p}),H} \longrightarrow \mathbf{M}_{0,H}$ is finite and flat.*

Proof (cf. [17], 5.1). We prove this result using the homogeneity principle, as in [17]. For this, we consider the set U in $\mathbf{M}_{0,H}$ consisting of those points x for which, at every lift to points $y \in \mathbf{M}_{U_1(\mathfrak{p}),H}$, the local ring at y is regular, and flat over the local ring at x. We prove that U has the following properties

(H1) U is open.
(H2) U contains all of $M_{0,H}$.
(H3) If U contains any ordinary point of $\mathbf{M}_{0,H}\otimes\kappa$, then it contains all such points.
(H4) If U contains any supersingular point of $\mathbf{M}_{0,H} \otimes \kappa$, then it contains all such points.
(H5) U contains a supersingular point of $\mathbf{M}_{0,H} \otimes \kappa$.

We will then conclude that $U = \mathbf{M}_{0,H}$.

(H1) is standard (see [17]) and (H2) follows from [3].

To prove (H3) and (H4), let x be a closed point of $\mathbf{M}_{0,H} \otimes \kappa$. If y is a closed point of $\mathbf{M}_{U_1(\mathfrak{p}),H}$ above x, the map $\mathcal{O}_{\mathbf{M}_{0,H},x} \longrightarrow \mathcal{O}_{\mathbf{M}_{U_1(\mathfrak{p}),H},y}$ is flat if and only if the induced map $\widehat{\mathcal{O}}^{sh}_{\mathbf{M}_{0,H},x} \longrightarrow \widehat{\mathcal{O}}^{sh}_{\mathbf{M}_{U_1(\mathfrak{p}),H},y}$ is flat.

Further, $\mathcal{O}_{\mathbf{M}_{U_1(\mathfrak{p}),H},y}$ is regular if and only if $\widehat{\mathcal{O}}^{sh}_{\mathbf{M}_{U_1(\mathfrak{p}),H},y}$ is regular. It thus suffices to consider the case where x is a geometric point of the special fibre.

Thus let $(\widehat{\mathbf{M}_{0,H}})_{(x)}$ denote the completion of the (strict) henselisation of $\mathbf{M}_{0,H}$ at x. Carayol ([3], 6.6) proves that $\mathbf{E}_\infty$ pulls back to this completion, and that

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

$\mathbf{E}_\infty|_{(\widehat{\mathbf{M}}_{0,H})_{(x)}}$ is the universal deformation of $\mathbf{E}_\infty|_x$. Thus the isomorphism class of the map $\widehat{\mathcal{O}}^{sh}_{\mathbf{M}_{0,H},x} \longrightarrow \widehat{\mathcal{O}}^{sh}_{\mathbf{M}_{U_1(\mathfrak{p}),H},y}$ depends only on the universal deformation of $\mathbf{E}_\infty|_x$, and thus only on whether x is an ordinary or supersingular point.

Finally we prove (H5).

Let x be a supersingular point of the special fibre of $\mathbf{M}_{0,H}$, and let y be a point of $\mathbf{M}_{U_1(\mathfrak{p}),H} \otimes \overline{\kappa}$ above x.

We conclude from the discussion above (see also [3], 7.4) that $(\widehat{\mathbf{M}}_{0,H})_{(x)} \xrightarrow{\sim} \mathrm{spec}(D_0^{\mathbf{E}_\infty|_x})$. To give a morphism $\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}) \longrightarrow \mathrm{Hom}_{\mathbf{M}_{0,H}-sch}((\widehat{\mathbf{M}}_{0,H})_{(x)}, \mathbf{E}_1)$ is the same as giving a map

$$\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}) \longrightarrow \mathrm{Hom}_{(\widehat{\mathbf{M}}_{0,H})_{(x)}-sch}((\widehat{\mathbf{M}}_{0,H})_{(x)}, \mathbf{E}_1|_{(\widehat{\mathbf{M}}_{0,H})_{(x)}}).$$

But this is equivalent to giving a map

$$\phi\colon (\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}) \longrightarrow \mathrm{Hom}_{(\widehat{\mathbf{M}}_{0,H})_{(x)}-sch}((\widehat{\mathbf{M}}_{0,H})_{(x)}, \mathbf{E}_\infty|_{(\widehat{\mathbf{M}}_{0,H})_{(x)}})$$

as the image of such a map must be killed by $\mathfrak{p}$. But this is exactly a $(\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p})$-map on the universal deformation. Thus to give such a morphism is to give a point of $\mathrm{spec}(L_1^{\mathbf{E}_\infty|_x})$.

But a $U_1(\mathfrak{p})$-structure on the universal deformation is exactly such a morphism with an extra condition which, by the results of Section 6, is closed. Thus there exists an ideal I of $L_1^{\mathbf{E}_\infty|_x}$ which expresses the condition that the image of the universal $(\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p})$-map defines an $\mathcal{O}_\mathfrak{p}$-subgroup of sections of the correct rank. Write $\widetilde{L}_1^{\mathbf{E}_\infty|_x} = L_1^{\mathbf{E}_\infty|_x}/I$. Then to give a $U_1(\mathfrak{p})$-structure on the universal deformation is to give a point of $\widetilde{L}_1^{\mathbf{E}_\infty|_x}$. In particular, $(\widehat{\mathbf{M}}_{U_1(\mathfrak{p}),H})_{(y)}$ may be identified with the closed subscheme $\mathrm{spec}(\widetilde{L}_1^{\mathbf{E}_\infty|_x})$ of $\mathrm{spec}(L_1^{\mathbf{E}_\infty|_x})$.

There is an obvious finite map $D_0^{\mathbf{E}_\infty|_x} \longrightarrow \widetilde{L}_1^{\mathbf{E}_\infty|_x}$. The induced map

$$\mathrm{spec}(\widetilde{L}_1^{\mathbf{E}_\infty|_x}) \longrightarrow \mathrm{spec}(D_0^{\mathbf{E}_\infty|_x})$$

is finite, so has closed image. But the map is a surjection when p is invertible, and so the image contains the (dense) set $\mathrm{spec}(D_0^{\mathbf{E}_\infty|_x} \otimes \mathbf{Z}[1/p])$. Thus the induced map is surjective.

But we saw earlier that for any supersingular $\mathcal{O}_\mathfrak{p}$-module E, the ring D_0^E is a regular ring of dimension 2. It follows that the (Krull) dimension of $\widetilde{L}_1^{\mathbf{E}_\infty|_x}$ is at least 2. But it is a quotient of the ring $L_1^{\mathbf{E}_\infty|_x}$, also regular of dimension 2, so that its maximal ideal is generated by 2 elements, and thus $\widetilde{L}_1^{\mathbf{E}_\infty|_x}$ is equal to $L_1^{\mathbf{E}_\infty|_x}$. But the morphism $D_0^{\mathbf{E}_\infty|_x} \longrightarrow L_1^{\mathbf{E}_\infty|_x}$ is flat, from Theorem 4.5.

This concludes the proof of (H5).

Finally, we observe that now $U = \mathbf{M}_{0,H}$. The set of supersingular points is finite and non-empty. U contains all supersingular points, and then, as U is open, it also contains some, and hence all, ordinary points. Thus it contains all of $M_{0,H}$ and all closed points in characteristic p. Then $U = \mathbf{M}_{0,H}$; if not, its complement would contain closed points not in U.

8. The $U_0(\mathfrak{p})$-Problem

We adopt an analogue of the definition of [17], 1.4 and 3.4.1.

DEFINITION 8.1. A $\mathfrak{p}$-*cyclic* subgroup scheme G of $\mathbf{E}_1|_S$ is one which, locally $fppf$ on S (i.e., after taking some faithfully flat, locally of finite presentation, morphism $T \longrightarrow S$), satisfies $G = \sum_{P\in\mathcal{K}}[P]$ for some set of sections $\mathcal{K}$ of the form $\langle P\rangle$, where P is a $U_1(\mathfrak{p})$-structure. We may say that G is $\mathfrak{p}$-*cyclic* when this occurs, and that P is a *generator* for G.

DEFINITION 8.2. A $U_0(\mathfrak{p})$-*structure* on S is a $\mathfrak{p}$-cyclic subgroup scheme G of $\mathbf{E}_1|_S$.

One defines a functor $\boldsymbol{\mathcal{M}}_{U_0(\mathfrak{p}),H}$ in the analogous way to that in which we defined the functor $\boldsymbol{\mathcal{M}}_{U_1(\mathfrak{p}),H}$, and once again we try to show that it is representable by exhibiting it as the closed subfunctor of a representable functor.

First, we confirm that if it is representable, it is indeed an integral model for $M_{U_0(\mathfrak{p}),H}$.

Define the functor $\mathcal{M}_{U_0(\mathfrak{p}),H}$ on $M_{0,H}$-schemes by setting $\mathcal{M}_{U_0(\mathfrak{p}),H}(S)$ to be the set of all subgroup schemes of $E_1|_S$ (étale) locally isomorphic to $(\mathcal{O}_\mathfrak{p}/\mathfrak{p})_S$.

LEMMA 8.3. *The two functors defined above coincide on $M_{0,H}$-schemes.*

Proof. To give an $\boldsymbol{\mathcal{M}}_{U_0(\mathfrak{p}),H}$-structure on an $M_{0,H}$-scheme S is to give a $\mathfrak{p}$-cyclic subgroup scheme of $\mathbf{E}_1|_S$. But $\mathbf{E}_1|_S = E_1|_S$, which is étale over S, so that $\mathfrak{p}$-cyclic subgroup schemes of $\mathbf{E}_1|_S$ must be locally isomorphic to $(\mathcal{O}_\mathfrak{p}/\mathfrak{p})_S$.

LEMMA 8.4. *The functor $\mathcal{M}_{U_0(\mathfrak{p}),H}$ is representable by the F-scheme $M_{U_0(\mathfrak{p}),H}$.*

Proof. In the same way as for the $U_1(\mathfrak{p})$-problem (Lemma 7.4), one observes that the equivalence classes of Drinfeld bases (P,Q) under the action of

$$\widetilde{U}_0(\mathfrak{p}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_\mathfrak{p}/\mathfrak{p}) \,\middle|\, c \in \mathfrak{p} \right\}$$

are clearly represented by the subgroup of sections $\langle P\rangle$, and thus by the subgroup schemes isomorphic to $(\mathcal{O}_\mathfrak{p}/\mathfrak{p})_S$. Then the quotient moduli problem (now identified with $\mathcal{M}_{U_0(\mathfrak{p}),H}$) will be represented by $M_{1,H}/\widetilde{U}_0(\mathfrak{p}) = M_{U_0(\mathfrak{p}),H}$.

Define the functor $\boldsymbol{\mathcal{M}}_{\mathfrak{p}-\mathrm{sub}}$ on $\mathbf{M}_{0,H}$-schemes by $\boldsymbol{\mathcal{M}}_{\mathfrak{p}-\mathrm{sub}}(S)$ is the set of finite flat $\mathcal{O}_\mathfrak{p}$-subgroup schemes of rank q of $\mathbf{E}_1|_S$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

PROPOSITION 8.5. *The functor* $\mathcal{M}_{\mathfrak{p}-\mathrm{sub}}$ *is representable by an* $\mathbf{M}_{0,H}$*-scheme* $\mathbf{M}_{\mathfrak{p}-\mathrm{sub}}$, *finite over* $\mathbf{M}_{0,H}$.

Proof. As in [17], 6.5.1, we regard $\mathbf{E}_1|_S$ as $\mathrm{spec}\,\mathcal{F}$, where $\mathcal{F}$ is a coherent sheaf of $\mathcal{O}_{(\mathfrak{p})}$-algebras on S which is locally free of rank q^2. A finite flat rank q subscheme of $\mathbf{E}_1|_S$ will correspond to a locally free rank q quotient of $\mathcal{F}$; the set of such quotients is represented by a Grassmannian. Then $\mathcal{M}_{\mathfrak{p}-\mathrm{sub}}$ is represented by the closed subscheme $\mathbf{M}_{\mathfrak{p}-\mathrm{sub}}$ of the Grassmannian over which the universal rank q quotient gives rise to an $\mathcal{O}_{\mathfrak{p}}$-subgroup scheme of the correct rank.

One proves that $\mathbf{M}_{\mathfrak{p}-\mathrm{sub}}$ is finite over $\mathbf{M}_{0,H}$ in the same way as [17]. We omit the details.

We would now like to observe that every $\mathcal{O}_{\mathfrak{p}}$-subgroup scheme of $\mathbf{E}_1|_S$ which is finite and locally free of rank q is a $\mathfrak{p}$-cyclic group, which is certainly true if p is invertible on S. For more general S, this seems difficult; in [17] it is derived as a corollary from the analogue of the main theorem below (whose proof is essentially identical to that of [17], 6.1.1).

If G is a $\mathfrak{p}$-cyclic group, then the functor on S-schemes

$$T \mapsto \text{generators of } G_T = G \times_S T_{/T}$$

is representable (as in [17], 1.10.13(1)) by a closed subscheme $G^\times$ of G.

Let $G \subset \mathbf{E}_1|_S$ be a $\mathfrak{p}$-cyclic subgroup. We work locally in the $fppf$ topology throughout, so that $G = \sum_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})}[\alpha P]$ for some section P of $\mathbf{E}_1$ over S. We fix such a generator P.

THEOREM 8.6. *In this situation,* $G^\times = \sum_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})^\times}[\alpha P]$.

Proof. Write $D = \sum_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})^\times}[\alpha P]$. Then $D \subset G$. The argument of [17], 6.1, continues to hold, so that we may deduce that there exists a closed immersion $D \hookrightarrow G^\times$.

To prove the equality of these schemes, we again follow [17].

Let $\mathbf{P}$ be the universal $U_1(\mathfrak{p})$-structure, and form

$$\mathbf{G} = \sum_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})}[\alpha\mathbf{P}] \quad \text{and} \quad \mathbf{D} = \sum_{\alpha\in(\mathcal{O}_{\mathfrak{p}}/\mathfrak{p})^\times}[\alpha\mathbf{P}].$$

Then define

$$\mathbf{M}_1 = \mathbf{M}_{U_1(\mathfrak{p}),H} \times_{\mathbf{M}_{0,H}} \mathbf{G}^\times, \quad \text{and} \quad \mathbf{M}_2 = \mathbf{M}_{U_1(\mathfrak{p}),H} \times_{\mathbf{M}_{0,H}} \mathbf{D}.$$

The morphism $\mathbf{D} \hookrightarrow \mathbf{G}^\times$ induces a morphism $\mathbf{M}_2 \longrightarrow \mathbf{M}_1$.

We prove that this morphism is an isomorphism using a variant of the homogeneity principle above.

Let U be the set of points x of $\mathbf{M}_{0,H}$ above which the map $\mathbf{M}_2 \longrightarrow \mathbf{M}_1$ is an isomorphism. To see that this is a sensible notion, and that U is open, we refer the reader to the proof of [17], 6.2. U contains all of $M_{0,H}$, as the result is clear if p is invertible.

For the analogues of (H3) and (H4) above, we observe that $\mathcal{O}_{\mathbf{M}_1,y_1}\xrightarrow{\sim}\mathcal{O}_{\mathbf{M}_2,y_2}$ if and only if $\widehat{\mathcal{O}}^{sh}_{\mathbf{M}_1,y_1}\xrightarrow{\sim}\widehat{\mathcal{O}}^{sh}_{\mathbf{M}_2,y_2}$ is an isomorphism. As before, this is a condition only on the universal deformation, and thus depends only on whether x is ordinary or supersingular.

Finally, we must prove that U contains the supersingular points. Thus, let x be such a point. To give a pair of $U_1(\mathfrak{p})$-structures on the universal deformation of $\mathbf{E}_\infty|_x$ is equivalent to giving a point of $R_1^{\mathbf{E}_\infty|_x} = L_1^{\mathbf{E}_\infty|_x} \otimes_{D_0^{\mathbf{E}_\infty|_x}} L_1^{\mathbf{E}_\infty|_x}$. Form the ring $M_1^{\mathbf{E}_\infty|_x} = R_1^{\mathbf{E}_\infty|_x}/I_M$ where we defined the ideal I_M during the proof of Theorem 5.2. The localisation of $\mathbf{M}_1$ above x is given by the scheme spec$(M_1^{\mathbf{E}_\infty|_x})$. The localisation of $\mathbf{M}_2$ is given by the scheme spec$(N_1^{\mathbf{E}_\infty|_x})$, where $N_1^{\mathbf{E}_\infty|_x} = R_1^{\mathbf{E}_\infty|_x}/I_N$, and I_N is the ideal

$$I_N = \left(\prod_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})^\times}(y_2 - f_\alpha(y_1))\right).$$

We claim that $I_N \subset I_M$. It suffices to show that the generator of I_N vanishes in $M_1^{\mathbf{E}_\infty|_x}$. But, in this ring:

$$\begin{aligned}\prod_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})^\times}(y_2 - f_\alpha(y_1)) &= \sum_{j=0}^{q-1} y_2^{q-j-1}\Sigma_j^{(q-1)}(\{f_\alpha(y_1)\}_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})^\times})\\ &= \sum_{j=0}^{q-1} y_2^{q-j-1}\Sigma_j^{(q)}(\{f_\alpha(y_1)\}_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})})\end{aligned}$$

(as $f_\mathfrak{p}(y_1) = 0$ in $M_1^{\mathbf{E}_\infty|_x}$)

$$= \sum_{j=0}^{q-1} y_2^{q-j-1}\Sigma_j^{(q)}(\{f_\alpha(y_2)\}_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})})$$

(by definition of I_M)

$$= \prod_{\alpha\in(\mathcal{O}_\mathfrak{p}/\mathfrak{p})^\times}(y_2 - f_\alpha(y_2))$$

(as above)

$$= 0 \qquad (\text{as } f_1(y_2) = y_2).$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

But the closed immersion $\mathbf{D} \hookrightarrow \mathbf{G}^{\times}$, gives rise to an inclusion $I_M \subset I_N$. Thus $I_M = I_N$, and the rings coincide, so that the analogue of (H5) follows.

Thus $D = G^{\times}$ as required.

We may now deduce an analogue of the 'Main Theorem on Cyclic Groups' ([17], 6.1.1). Indeed, Theorem 8.6 is exactly the analogue of the second half of this theorem. For the first half, we want to show that if G is an $\mathcal{O}_{\mathfrak{p}}$-subgroup scheme of $\mathbf{E}_1|_S$, which is finite and locally free of rank q, then G is $\mathfrak{p}$-cyclic if and only if $G^{\times}$ is finite and locally free of rank $q-1$. Certainly, if G is $\mathfrak{p}$-cyclic, $G^{\times}$ is a finite group scheme; it is also clearly flat, as Theorem 8.6 exhibits it as a closed subscheme of G, which is flat. Its rank may be computed after inverting p, when the result is clear. Conversely, if G is not $\mathfrak{p}$-cyclic, its scheme of generators has no field-valued points.

PROPOSITION 8.7. *Let $G \subset \mathbf{E}_1|_S$ be a finite flat $\mathcal{O}_{\mathfrak{p}}$-subgroup scheme of rank q. There exists a closed subscheme $S_{\mathfrak{p}-\mathrm{cyclic}}$ of S such that, if S' is an S-scheme, then $G \times_S S'$ is $\mathfrak{p}$-cyclic if and only if $S' \longrightarrow S$ factors through $S_{\mathfrak{p}-\mathrm{cyclic}}$.*

Proof. This is a trivial variant of [17], Section 6.4, given Theorem 8.6.

LEMMA 8.8. *$\underline{\boldsymbol{\mathcal{M}}}_{U_0(\mathfrak{p}),H}$ is representable by an $\mathbf{M}_{0,H}$-scheme $\mathbf{M}_{U_0(\mathfrak{p}),H}$, finite over $\mathbf{M}_{0,H}$.*

Proof. Simply because now $\boldsymbol{\mathcal{M}}_{U_0(\mathfrak{p}),H}$ is a closed subfunctor of the functor $\boldsymbol{\mathcal{M}}_{\mathfrak{p}-\mathrm{sub}}$, which we proved was representable by a finite $\mathbf{M}_{0,H}$-scheme $\mathbf{M}_{\mathfrak{p}-\mathrm{sub}}$. The result follows.

THEOREM 8.9. *The scheme $\mathbf{M}_{U_0(\mathfrak{p}),H}$ is a regular scheme (of dimension 2). The projection morphism $\mathbf{M}_{U_0(\mathfrak{p}),H} \longrightarrow \mathbf{M}_{0,H}$ is finite and flat.*

Proof. This follows in the same way as [17], 6.6.1.

Remark. In the sequel, we only consider the schemes $\mathbf{M}_{U_0(\mathfrak{p}),H}$. The same methods as those above suffice also to prove the existence of integral models $\mathbf{M}_{U_1(\mathfrak{p}^n),H}$ and $\mathbf{M}_{U_0(\mathfrak{p}^n),H}$. The proofs are the essentially identical to those above. The ring L_1^E of Theorem 4.5 should be replaced by

$$L_n^E = D_0^E[[y_1]]/(f_{\mathfrak{p}^n}(y_1)/f_{\mathfrak{p}^{n-1}}(y_1)),$$

which again has $\{y_1, t_1\}$ as a regular sequence of parameters. The only substantial change required in any of the proofs occurs at the end of the proof of Theorem 8.6; the calculation to show that $I_N \subset I_M$ seems rather harder for $n > 1$, but one can adapt the methods of [17], Section 6.3, to this situation.

9. $\mathfrak{p}$-Cyclic Isogenies

This section closely follows parts of [17], Ch.13.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

DEFINITION 9.1. If Z is a scheme over $S = \operatorname{spec}\overline{\kappa}$, and $\tau \in \operatorname{Gal}(\overline{\kappa}/\kappa)$, we can form $Z^{(\tau)} = Z \otimes_{\overline{\kappa}\nearrow_\tau} \overline{\kappa}$.

Let σ denote the arithmetic Frobenius element, then $Z^{(\sigma)}$ is the image of the relative Frobenius map $F_{Z/\overline{\kappa}}$ on Z. In this case, and more generally when S is the spectrum of any perfect κ-algebra, we can form $Z^{(\sigma^{-1})}$. We will write F for the relative Frobenius morphism, and will denote the absolute Frobenius morphism by $F_{\text{abs}}: S \longrightarrow S$, defined, for S a κ-scheme, by the map $s \mapsto s^q$ on the affine rings.

If x is a geometric point of $\mathbf{M}_{U_0(\mathfrak{p}),H}$, we say that a $U_0(\mathfrak{p})$-structure at x is a $\mathfrak{p}$-cyclic subgroup scheme $G \subset \mathbf{E}_1|_x \subset \mathbf{E}_\infty|_x$.

DEFINITION 9.2. By a $\mathfrak{p}$-*cyclic isogeny*, we will mean an $\mathcal{O}_\mathfrak{p}$-linear isogeny between divisible $\mathcal{O}_\mathfrak{p}$-modules whose kernel is a $\mathfrak{p}$-cyclic subgroup scheme.

If G is a $U_0(\mathfrak{p})$-structure at x, we may associate to it the $\mathfrak{p}$-cyclic isogeny $\mathbf{E}_\infty|_x \twoheadrightarrow \mathbf{E}_\infty|_x/G$. We will study possible isogenies of this type, and use this to deduce that there is a very similar description of $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \kappa$ as that which exists for the corresponding modular curves.

Recall that if x is a closed point of $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \kappa$ then it is either ordinary or supersingular. We shall show that if x is supersingular, then there is exactly one $U_0(\mathfrak{p})$-structure on $\mathbf{E}_\infty|_x$, and if x is ordinary, then there are exactly two, corresponding to the kernel of the Frobenius morphism F, and to the kernel of the Verschiebung, which we denote by V.

As is well known, the kernel of the action of Frobenius on the divisible $\mathcal{O}_\mathfrak{p}$-modules is concentrated in the connected part.

If x is supersingular, then $\mathbf{E}_\infty|_x$ is connected, so that $\mathbf{E}_\infty|_x(\overline{\kappa}) = \{0\}$, and thus the only possible $U_0(\mathfrak{p})$-structure at x is that defined by the ideal sheaf which is the product of q copies of the ideal sheaf defining the zero section. This is an $\mathcal{O}_\mathfrak{p}$-subgroup scheme, as it coincides with the kernel of the Frobenius morphism. (Note that we are using here the fact that $\mathbf{E}_\infty|_x$ is a p-divisible group of dimension 1; the analogous result would be false for p-divisible groups of higher dimension.)

We concentrate for the remainder of this section on the case where x is ordinary.

In this case, one has a connected-étale decomposition of $\mathbf{E}_\infty|_x$ as follows:

$$0 \longrightarrow \mathbf{L}_\infty|_x \longrightarrow \mathbf{E}_\infty|_x \longrightarrow \mathrm{F}_\mathfrak{p}/\mathcal{O}_\mathfrak{p} \longrightarrow 0.$$

Write $\mathbf{L}_n|_x$ (resp. $\mathbf{E}_n|_x$) for the $\mathfrak{p}^n$-torsion points of $\mathbf{L}_\infty|_x$ (resp. $\mathbf{E}_\infty|_x$).

PROPOSITION 9.3. *In the diagram*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbf{L}_1|_x & \longrightarrow & \mathbf{E}_1|_x & \longrightarrow & \mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p} & \longrightarrow & 0 \\ & & & & \uparrow & \nearrow \phi & & & \\ & & & & G & & & & \end{array}$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

where $G \subset \mathbf{E}_1|_x$ *is a* $U_0(\mathfrak{p})$*-structure at* x*, we have either*

(1) $\operatorname{im}\phi = \mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$, *and* $G \cap \mathbf{L}_1|_x = \{0\}$, *or*
(2) $\operatorname{im}\phi = \{0\}$ *and* $G = \mathbf{L}_1|_x$.

Proof. $\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$ is étale, so that $\phi(G)$ is a finite étale $\mathcal{O}_\mathfrak{p}$-subgroup of $\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$. Thus $\phi(G) = \{0\}$ or is $\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$ (if one has a nontrivial section, its translates under $\mathcal{O}_\mathfrak{p}$ give all of $\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$). In the first case, one sees immediately that $G \subset \mathbf{L}_1|_x$; by comparing ranks, one deduces an equality. In the second case, we see that ϕ is an isomorphism, and thus that $G \cap \mathbf{L}_1|_x = \ker\phi = \{0\}$.

DEFINITION 9.4. Recall that we have a map $F\colon \mathbf{E}_\infty|_x \longrightarrow \mathbf{E}_\infty|_x^{(\sigma)}$. We define the *Verschiebung* V to be the Cartier dual of the map

$$F\colon (\mathbf{E}_\infty|_x)^D \longrightarrow (\mathbf{E}_\infty|_x^D)^{(\sigma)} = (\mathbf{E}_\infty|_x^{(\sigma)})^D,$$

where $(-)^D$ denotes Cartier duality. One knows that $\ker F$ is connected. By comparing ranks, it is clear that $\ker F = \mathbf{L}_1|_x$.

Also, $\ker V$ is Cartier dual to the kernel of F (applied to $\mathbf{E}_\infty|_x^D$), so that $\ker V$ is étale and contained in the $\mathfrak{p}$-torsion of $\mathrm{F}_\mathfrak{p}/\mathcal{O}_\mathfrak{p}$, so that (comparing ranks again) $\ker V$ is (a possibly twisted form of) $\mathfrak{p}^{-1}/\mathcal{O}_\mathfrak{p}$.

LEMMA 9.5. *Let* x *be a geometric point of* $\mathbf{M}_0 \otimes \kappa$*. Then* $(\mathbf{E}_\infty|_x)^{(\sigma)} \cong \mathbf{E}_\infty|_{F_{\text{abs}}x}$.

Proof. For $x\colon \operatorname{spec}\overline{\kappa} \longrightarrow \mathbf{M}_0 \otimes \kappa$, one has the following Cartesian diagrams:

$$\begin{array}{ccccc}
\mathbf{E}_\infty^{(q)}|_x & \longrightarrow & \mathbf{E}_\infty^{(q)} & \longrightarrow & \mathbf{E}_\infty \\
\downarrow & & \downarrow & & \downarrow \\
\operatorname{spec}\overline{\kappa} & \xrightarrow{x} & \mathbf{M}_0 \otimes \kappa & \xrightarrow{F_{\text{abs}}} & \mathbf{M}_0 \otimes \kappa,
\end{array}$$

where $\mathbf{E}_\infty^{(q)}$ is the pull-back in the right-hand square, and

$$\begin{array}{ccccc}
\mathbf{E}_\infty|_x^{(\sigma)} & \longrightarrow & \mathbf{E}_\infty|_x & \longrightarrow & \mathbf{E}_\infty \\
\downarrow & & \downarrow & & \downarrow \\
\operatorname{spec}\overline{\kappa} & \xrightarrow{F_{\text{abs}}} & \operatorname{spec}\overline{\kappa} & \xrightarrow{x} & \mathbf{M}_0 \otimes \kappa.
\end{array}$$

But one knows that there is a commutative diagram

$$\begin{array}{ccc}
\operatorname{spec}\overline{\kappa} & \xrightarrow{F_{\text{abs}}} & \operatorname{spec}\overline{\kappa} \\
\downarrow x & & \downarrow x \\
\mathbf{M}_0 \otimes \kappa & \xrightarrow{F_{\text{abs}}} & \mathbf{M}_0 \otimes \kappa
\end{array}$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

so that one deduces a natural isomorphism $\mathbf{E}_\infty^{(q)}|_x \xrightarrow{\sim} \mathbf{E}_\infty|_x^{(\sigma)}$, both being pull-backs in the same Cartesian square. Thus the first diagram may be rewritten

$$\begin{array}{ccc} (\mathbf{E}_\infty|_x)^{(\sigma)} & \longrightarrow & \mathbf{E}_\infty \\ \downarrow & & \downarrow \\ \operatorname{spec}\overline{\kappa} & \xrightarrow{F_{\text{abs}}\circ x} & \mathbf{M}_0 \otimes \kappa, \end{array}$$

and so $(\mathbf{E}_\infty|_x)^{(\sigma)}$ is identified with $\mathbf{E}_\infty|_{F_{\text{abs}}x}$, defined as the pull-back in the last diagram above.

COROLLARY 9.6. *Let* $x\colon \operatorname{spec}\overline{\kappa} \longrightarrow \mathbf{M}_{0,H}\otimes\kappa$ *be a geometric point of* $\mathbf{M}_{0,H}\otimes\kappa$. *Then* $(\mathbf{E}_\infty|_x)^{(\sigma)} \cong \mathbf{E}_\infty|_{F_{\text{abs}}x}$.

Proof. If y is a geometric point of $\mathbf{M}_0\otimes\kappa$ above x, then $\mathbf{E}_\infty|_x = \mathbf{E}_\infty|_y$. The result now follows from the previous lemma.

PROPOSITION 9.7. *A* $\mathfrak{p}$*-cyclic isogeny* π *from* $\mathbf{E}_\infty|_x$ *is, up to isomorphism, either*

$$\pi : \mathbf{E}_\infty|_x \xrightarrow{F} \mathbf{E}_\infty|_{F_{\text{abs}}x} \quad \textit{or} \quad \pi : \mathbf{E}_\infty|_x \xrightarrow{V} \mathbf{E}_\infty|_y,$$

where $F_{\text{abs}}\, y = x$.

Proof. If $G = \ker\pi$, then, by the previous results, either $G = \ker F$, which clearly gives the first possibility; otherwise, $G^D = \ker F$, and on dualising again, we recover the second possibility.

10. The Structure of the Special Fibre of $\mathbf{M}_{U_0(\mathfrak{p}),H}$

To describe the special fibre of $\mathbf{M}_{0,H}$, we make use of the crossings theorem of Katz–Mazur [17], 13.1.3.

We first check that the hypotheses hold.

PROPOSITION 10.1. *The complete local ring of* $\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$ *at points lying above supersingular points of* $\mathbf{M}_{0,H}\otimes\kappa$ *is isomorphic to* $\kappa[[X,Y]]/(\textit{one equation})$.

Proof. This follows from a characterisation of complete local rings (see, for instance, [19]), and follows exactly as in [17], 13.2.

We can use the crossings theorem [17] to conclude that $\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$ consists of two copies of $\mathbf{M}_{0,H}\otimes\kappa$ which intersect transversally above each supersingular point of $\mathbf{M}_{0,H}\otimes\kappa$ (a set which is explicitly described in [3]). In order to use the crossings theorem, we prove the following result:

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

THEOREM 10.2. *There exist two κ-schemes, $Z_{0,H}$ and $Z_{1,H}$, with*

$$\begin{array}{ccc} \mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa & \longleftarrow & Z_{0,H}\sqcup Z_{1,H} \\ \downarrow & \swarrow & \\ \mathbf{M}_{0,H}\otimes\kappa & & \end{array}$$

such that

(1) *There is a unique closed point of each $Z_{i,H}$ above each supersingular point of $\mathbf{M}_{0,H}\otimes\kappa$.*
(2) *Each $Z_{i,H}$ is finite and flat over $\mathbf{M}_{0,H}\otimes\kappa$.*
(3) *Each $(Z_{i,H})^{\mathrm{red}}$ is a smooth curve over κ.*
(4) *$Z_{i,H}\longrightarrow\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$ is a closed immersion,*

and the union

$$Z_{0,H}\sqcup Z_{1,H}\longrightarrow\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$$

is an isomorphism away from supersingular points of $\mathbf{M}_{0,H}\otimes\kappa$.

Proof. The proof of this result is exactly analogous to the proof of [17], 13.4.4.

To give a $\mathfrak{p}$-cyclic isogeny at an ordinary point x is to give one of the following morphisms (Propositon 9.7):

$$F\colon\ \mathbf{E}_\infty|_x\longrightarrow\mathbf{E}_\infty|_{x\circ\sigma}\quad\text{or}\quad V\colon\ \mathbf{E}_\infty|_x\longrightarrow\mathbf{E}_\infty|_{x\circ\sigma^{-1}}.$$

Define both $Z_{0,H}$ and $Z_{1,H}$ to be copies of $\mathbf{M}_{0,H}\otimes\kappa$. A geometric point x of $Z_{0,H}$ will correspond to the $U_0(\mathfrak{p})$-structure $\ker F$ at x, and the geometric point x of $Z_{1,H}$ will correspond to the $U_0(\mathfrak{p})$-structure $\ker V$ at x. These coincide precisely above the supersingular points of $\mathbf{M}_{0,H}\otimes\kappa$. The maps $Z_{i,H}\longrightarrow\mathbf{M}_{0,H}\otimes\kappa$ are finite and flat, as they are surjective morphisms between smooth curves.

In order to see that there is a map $Z_{0,H}\longrightarrow\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$, it is necessary to verify that the kernel of the relative Frobenius $\mathbf{E}_1|_{\mathbf{M}_{0,H}\otimes\kappa}\longrightarrow\mathbf{E}_1|^{(q)}_{\mathbf{M}_{0,H}\otimes\kappa}$ is $\mathfrak{p}$-cyclic. But this kernel is a closed subgroup of $\mathbf{E}_1|_{\mathbf{M}_{0,H}\otimes\kappa}$, which is flat over $\mathbf{M}_{0,H}\otimes\kappa$, and so it is flat. Finiteness is trivial. Its rank may be computed locally, at ordinary points say, as in Section 9. Finally, as the kernel is connected, 0 is a generator. To get a map $Z_{1,H}\longrightarrow\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$, it is necessary to verify that the kernel of the Verschiebung is $\mathfrak{p}$-cyclic. But it is Cartier dual to Frobenius, and so it suffices to check that the Cartier dual of a $\mathfrak{p}$-cyclic group scheme is again $\mathfrak{p}$-cyclic. For this, one imitates [17], Section 5.5.

The remainder of the proof is exactly the same as [17], 13.4.4.

It follows that $\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\kappa$ consists of two copies of $\mathbf{M}_{0,H}\otimes\kappa$, intersecting transversally above the finite set of supersingular points.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

Note 10.3. We have been particularly concerned with the $U_0(\mathfrak{p})$-problem in this section, as it is this analysis which is used in the work of Ribet and others on levels for modular mod ℓ representations, but it is now also very easy to give the structure of $\mathbf{M}_{U_1(\mathfrak{p}),H} \otimes \kappa$. The reader will have no problems with extending the analysis of [17], Section 13.5 to this case. In particular, the special fibre will consist of two parts; the first is essentially a copy of $\mathbf{M}_{0,H} \otimes \kappa$ (suitably thickened, which will represent the generators for $\ker F$), and the second is an Igusa curve, the scheme of generators for $\ker V$.

11. Galois Representations

In this section, we begin the study of Galois representations.

DEFINITION 11.1. Let $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ be a continuous irreducible semisimple representation. We say $\overline{\rho}$ is *modular* if there exists a Hilbert cuspidal eigenform f, and a prime λ above ℓ of the number field generated by the Hecke eigenvalues of f, such that $\overline{\rho}$ is the semisimplification of the reduction of the λ-adic representation associated to f. We may also say that $\overline{\rho}$ is *attached* to f when this occurs.

CONJECTURE 11.2. Let $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ be a continuous irreducible semisimple representation. Then if $\overline{\rho}$ is modular, it is attached to some Hilbert cuspidal eigenform f_0 whose level is equal to the Artin conductor in characteristic ℓ of $\overline{\rho}$ (see [24], 1.2).

This conjecture is a generalisation of part of the strong Serre conjectures to totally real fields. This part of the Serre conjectures was proven by Ribet [22] and Diamond [10] in the case where $\mathrm{F} = \mathbf{Q}$ (at least when ℓ is odd); the earliest part of the proof is due to Mazur, and is known as Mazur's Principle. In this section, we exploit the structure of the Shimura curves that we have studied to prove an analogue of Mazur's Principle when F is a totally real field of odd degree over $\mathbf{Q}$. The proof nearly works for extensions of even degree also, but there is a case which this method does not address.

We will prove the following theorem in Sections 13–18, and will explain in the next section how to deduce our version of Mazur's Principle.

THEOREM 11.3. *Assume* $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$, *a continuous irreducible semisimple representation, is attached to a Hilbert cuspidal eigenform* $f \in S_{k,w}(U_0(\mathfrak{p}) \cap U)$, *where* $\mathfrak{p} \nmid \ell$, $k \geqslant 2t$, *and* U *is sufficiently small. Suppose* U *decomposes as* $\prod_{\mathfrak{q}} U_{\mathfrak{q}}$, *with* $U_{\mathfrak{p}} = \mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}})$. *If* $[\mathrm{F} : \mathbf{Q}]$ *is even, suppose also that there exists some finite place* $\mathfrak{q}_0 \neq \mathfrak{p}$ *of* F *at which the automorphic representation corresponding to* f *is special or supercuspidal. Then if* $\overline{\rho}$ *is irreducible, and unramified at* $\mathfrak{p}$, *and* $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$, *there exists a Hilbert cuspidal eigenform* $f' \in S_{k,w}(U)$ *to which* $\overline{\rho}$ *is attached.*

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

We fix once and for all some $\overline{\rho}$: $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ which satisfies the conditions of the theorem.

We now make a choice of a quaternion algebra B for the rest of the paper. If $[\mathrm{F} : \mathbf{Q}]$ is odd, we choose B to be the quaternion algebra ramified exactly at $\tau_2, \ldots, \tau_d$, and at no other places. If $[\mathrm{F} : \mathbf{Q}]$ is even, we choose B to be the quaternion algebra ramified at the infinite places $\tau_2, \ldots, \tau_d$ and at the finite place $\mathfrak{q}_0$. Then, in both of these cases, the Jacquet–Langlands correspondence provides an automorphic representation π on B corresponding to f.

Fix, for every place v at which B is split, an isomorphism

$$(\mathcal{O}_B \otimes_{\mathcal{O}_\mathrm{F}} \mathcal{O}_{\mathrm{F},v})^\times \xrightarrow{\sim} \mathrm{GL}_2(\mathcal{O}_{\mathrm{F},v}).$$

We write G for the $\mathbf{Q}$-algebraic group $\mathrm{Res}_{\mathrm{F}/\mathbf{Q}}(B^\times)$, and denote its centre by Z.

Then $(\pi^\infty)^{H.U_0(\mathfrak{p})}$ will be non-trivial for some open compact subgroup H (corresponding to U) of the restricted product, taken over all finite places $v \neq \mathfrak{p}$ of F, of the groups $(B \otimes \mathrm{F}_v)^\times$. When U is sufficiently small, H will be sufficiently small so that $\mathbf{E}_1$ exists on $\mathbf{M}_{0,H}$. In this case, we have demonstrated the existence of an integral model for the curve $M_{U_0(\mathfrak{p}),H}$, and thus we can consider the reduction of this model in characteristic p (as above).

In the next section, under an additional hypothesis on ℓ, we will explain how to adapt a trick of Diamond and Taylor (see [11], Lemma 11, [12], Lemma 3 and [7], Lemma 4.11) to replace the hypothesis that U be sufficiently small with a hypothesis on ℓ. In particular, we will be able to reformulate our main result in terms of the more familiar groups $U_1(\mathfrak{n})$.

We write (following Carayol [4]) C for the set of cuspidal automorphic representations on B of weight (k, w). Write $S_{k,w}(H \cap U_0(\mathfrak{p}))^B$ for the direct sum $\bigoplus_{\pi \in C} (\pi^\infty)^{H.U_0(\mathfrak{p})}$, and $S_{k,w}(H)^B$ for $\bigoplus_{\pi \in C} (\pi^\infty)^{H.\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})}$.

We compute cohomology groups of some of the Shimura curves that we studied in the previous sections. It is in (analogues of) these cohomology groups that Carayol originally found representations associated to Hilbert modular forms, and by using the tools of cohomology theory, we will be able to analyse the representations that Carayol constructs in some detail, in a similar way to [4].

Most of this study will mimic that of Carayol [4], although with a different Shimura curve, and a different sheaf.

12. Mazur's Principle

In this section, we will remove the hypothesis that U be sufficiently small, and show how Theorem 11.3 implies the following theorem:

THEOREM (Mazur's Principle). *Assume* $\overline{\rho}$: $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$, *a continuous irreducible semisimple representation, is attached to a Hilbert cuspidal eigenform* $f \in S_{k,w}(U_0(\mathfrak{p}) \cap U_1(\mathfrak{n}))$, *where* $\mathfrak{p} \nmid \mathfrak{n}\ell$, *and* $k \geqslant 2t$. *Suppose* $[\mathrm{F}(\mu_\ell) : \mathrm{F}] \geqslant 4$.

If $\mathrm{F}/\mathbf{Q}$ *has even degree, suppose also that there exists some finite place* $\mathfrak{q}_0 \neq \mathfrak{p}$ *of* F *at which the automorphic representation corresponding to* f *is special or supercuspidal. Then if* $\overline{\rho}$ *is irreducible, and unramified at* $\mathfrak{p}$*, and* $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$*, there exists a Hilbert cuspidal eigenform* $f' \in S_{k,w}(U_1(\mathfrak{n}))$ *to which* $\overline{\rho}$ *is attached.*

Recall that we have proven the existence of integral models for the Shimura curves in which we are interested under the assumption that H is sufficiently small that $\mathbf{E}_1$ exists on $\mathbf{M}_{0,H}$, an apparently stronger condition ([3], 1.4.1.2) than that H be sufficiently small that $\mathbf{M}_{0,H}$ exists. The existence of $\mathbf{M}_{0,H}$ follows from combining the conditions [3] 1.4.1.1, 4.5.2 and 5.3. We begin by finding an effective criterion for the existence of the integral models.

LEMMA 12.1. $\mathbf{M}_{0,H}$ *exists if* H *is contained in a subgroup of the form*

$$U_1^1(\mathfrak{n}) = \left\{\alpha \in G(\mathbf{A}^\infty) \,\middle|\, \alpha \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{\mathfrak{n}}\right\}$$

if $\mathfrak{n}$ *is an ideal of* $\mathcal{O}_\mathrm{F}$ *satisfying* $\mathfrak{p} \nmid \mathfrak{n}$ *and* $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{n}) > 4^d$.

Proof. We first show that, under the hypothesis on H, the condition [3], 1.4.1.1 is satisfied.

But with the notation of the proof of [3] 1.4.1.1, one finds, for each embedding $\tau\colon \mathrm{F} \hookrightarrow \mathbf{R}$, $-4 \leqslant \tau(\mu) \leqslant 0$.

On the other hand, $\mu \in \mathfrak{n}$. Under the hypothesis on $\mathfrak{n}$, one concludes that $\mu = 0$, as required.

Let H' be a normal subgroup of H of finite index, sufficiently small that [3] 4.5.2 and 5.3 hold (in fact, 5.3 is implied by the second assumption on H [16] and 4.5.2 is satisfied by some normal subgroup of finite index, using a result of Chevalley [6]). The Galois group of the covering $M_{0,H'} \longrightarrow M_{0,H}$ is

$$H_{H'} = H/H'(Z(\mathbf{Q}) \cap H.\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})).$$

$H_{H'}$ acts on $\mathbf{M}_{0,H'}$ by [3], 6.2. Define $\mathbf{M}_{0,H} = \mathbf{M}_{0,H'}/H_{H'}$, which exists as a quasiprojective scheme over $\operatorname{spec} \mathcal{O}_{(\mathfrak{p})}$ by [14].

Because the action of $H_{H'}$ on $\mathbf{M}_{0,H'}$ is free (this follows from [3], 1.4.1.1 and 6.2), the quotient map $q\colon \mathbf{M}_{0,H'} \longrightarrow \mathbf{M}_{0,H}$ is étale, so that $\mathbf{M}_{0,H}$ is also smooth over $\operatorname{spec} \mathcal{O}_{(\mathfrak{p})}$. Further, $\mathbf{M}_{0,H'}$ is proper over $\operatorname{spec} \mathcal{O}_{(\mathfrak{p})}$. We verify the valuative criterion for properness for the scheme $\mathbf{M}_{0,H}$. If R is a discrete valuation $\mathcal{O}_{(\mathfrak{p})}$-algebra, with field of fractions K, then given a K-valued point x of $\mathbf{M}_{0,H}$, we may lift it to a point $y\colon \operatorname{spec} K \longrightarrow \mathbf{M}_{0,H'}$, and by properness, there is a unique extension of y to a map $\widetilde{y}\colon \operatorname{spec} R \longrightarrow \mathbf{M}_{0,H'}$, which induces (after composing with q) a point $\widetilde{x}\colon \operatorname{spec} R \longrightarrow \mathbf{M}_{0,H}$. Any two lifts of x differ by an element of the Galois group of the covering, and one immediately verifies that the induced point $\widetilde{x}$ is independent of the choice of lift. Thus $\widetilde{x}$ is unique, and properness follows.

Remark. The same conclusion holds if H is contained in

$$U(\mathfrak{n}) = \left\{ \alpha \in G(\mathbf{A}^\infty) \,\middle|\, \alpha \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{\mathfrak{n}} \right\}$$

if $\mathfrak{p} \nmid \mathfrak{n}$ and $N_{F/\mathbf{Q}}(\mathfrak{n}) > 2^d$. The same proof works as in Lemma 12.1, except that one knows now that $\mu \in \mathfrak{n}^2$.

LEMMA 12.2. *Suppose H satisfies the hypothesis of Lemma* 12.1. *Then there is an integral model for $\mathbf{M}_{U_0(\mathfrak{p}),H}$ whose special fibre consists of two copies of $\mathbf{M}_{0,H} \otimes \kappa$ intersecting transversally above supersingular points.*

Proof. Choose a normal subgroup H' of H of finite index which is sufficiently small that $\mathbf{E}_1$ exists on $\mathbf{M}_{0,H'}$. The Galois group of the covering $M_{U_0(\mathfrak{p}),H'} \longrightarrow M_{U_0(\mathfrak{p}),H}$ is

$$H_{H'} = H/H'(Z(\mathbf{Q}) \cap H.U_0(\mathfrak{p})).$$

But $Z(\mathbf{Q}) \cap H.U_0(\mathfrak{p}) = Z(\mathbf{Q}) \cap H.\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})$, so that

$$H_{H'} = H/H'(Z(\mathbf{Q}) \cap H.\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})).$$

$H_{H'}$ acts freely on $\mathbf{M}_{0,H'}$ (Lemma 12.1 and [3], 1.4.1.1), and therefore it acts freely on $\mathbf{M}_{U_0(\mathfrak{p}),H'}$. Then we define $\mathbf{M}_{U_0(\mathfrak{p}),H} = \mathbf{M}_{U_0(\mathfrak{p}),H'}/H_{H'}$, the quotient scheme. This exists as the action of $H_{H'}$ is free, and the quotient map is finite and étale ([SGA 1], V2.3). It follows that formation of the quotient scheme commutes with base change, so that one deduces the same description of the special fibre, consisting of two copies of $\mathbf{M}_{0,H} \otimes \kappa$ intersecting transversally above the supersingular points, as in Section 10, even under this weaker assumption on H.

Remark. In both Lemmas 12.1 and 12.2, we defined the integral models with reference to a suitably chosen smaller subgroup of H. In fact, the resulting model is, up to unique isomorphism, independent of the choice made. For if H'' is a second subgroup, we may assume without loss of generality that $H'' \subset H'$ (otherwise consider $H' \cap H''$), and note that one has canonically an isomorphism $\mathbf{M}_{0,H'} = \mathbf{M}_{0,H''}/H'_{H''}$. A similar line of reasoning holds for Lemma 12.2. Exactly as in [3], 6.2, we may conclude that the collections $\{\mathbf{M}_{0,H}\}$ and $\{\mathbf{M}_{U_0(\mathfrak{p}),H}\}$ form projective systems, with finite étale transition maps, as H varies.

Under a supplementary condition on F and ℓ, we now show that we may also remove the hypothesis that U (equivalently, H) is sufficiently small in Theorem 11.3. To do this, we introduce some auxiliary level structure to ensure that H is sufficiently small, and then remove this extra structure at the end of the proof.

Suppose now that the given mod ℓ representation $\overline{\rho}$ has level $U_1(\mathfrak{n}) \cap U_0(\mathfrak{p})$. In other words, there exists a cuspidal eigenform of level $U_1(\mathfrak{n}) \cap U_0(\mathfrak{p})$ giving rise to $\overline{\rho}$.

We now copy [11] to find an auxiliary prime $\mathfrak{q}_1$ such that

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

- $N_{F/\mathbf{Q}}(\mathfrak{q}_1) > 4^d$,
- $\mathfrak{q}_1 \nmid \mathfrak{n}\ell\mathfrak{p}$,
- there are no congruences between forms of level $U_1(\mathfrak{n})$ and $\mathfrak{q}_1$-new forms of level dividing $U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{q}_1)$.

Then, throughout the proof of the main theorem above, we work with an auxiliary $U_1^1(\mathfrak{q}_1)$-level structure to remove $U_0(\mathfrak{p})$ from the level, and finally we may remove the auxiliary level structure, as we have chosen it in such a way that there are no congruences.

One can classify the cases in which congruences occur, by means of a generalisation of a result of Carayol ([5]). One knows that if two cuspidal automorphic representations π and π' of level $\mathfrak{m}$ and $\mathfrak{n}$ are congruent mod ℓ, then $v_{\mathfrak{q}}(\mathfrak{m}) = v_{\mathfrak{q}}(\mathfrak{n})$ (for $\mathfrak{q} \nmid \ell$) unless

- $N_{F/\mathbf{Q}}(\mathfrak{q}) \equiv -1 \pmod{\ell}$, and one of $\pi_{\mathfrak{q}}$ and $\pi'_{\mathfrak{q}}$ is a supercuspidal Weil representation, or
- $N_{F/\mathbf{Q}}(\mathfrak{q}) \equiv 1 \pmod{\ell}$, or
- one of $\pi_{\mathfrak{q}}$ and $\pi'_{\mathfrak{q}}$ is special unramified, so that, if $\overline{\rho}$ denotes the corresponding mod ℓ Galois representation,

$$\operatorname{tr}\overline{\rho}(\mathsf{Frob}_{\mathfrak{q}})^2/\det\overline{\rho}(\mathsf{Frob}_{\mathfrak{q}}) = (1 + N_{F/\mathbf{Q}}(\mathfrak{q}))^2/N_{F/\mathbf{Q}}(\mathfrak{q}).$$

We first note that if a representation $\pi_{\mathfrak{q}}$ of $\mathrm{GL}_2(F_{\mathfrak{q}})$ has a fixed vector under $U_1^1(\mathfrak{q})$, then it cannot be supercuspidal. For this, one considers the action of the abelian group $U_1(\mathfrak{q})/U_1^1(\mathfrak{q})$ on the $U_1^1(\mathfrak{q})$-fixed vectors of $\pi_{\mathfrak{q}}$. This action decomposes into characters; twisting by one such gives a representation with a vector fixed under $U_1(\mathfrak{q})$, which cannot be supercuspidal. Thus a twist of $\pi_{\mathfrak{q}}$ is not supercuspidal, which implies the same result for $\pi_{\mathfrak{q}}$. We thank Fred Diamond for the above observation.

Thus congruences between forms of level $U_1(\mathfrak{n})$ and forms of level $U_1(\mathfrak{n})\cap U_1^1(\mathfrak{q})$ can only occur in the latter two of the above classification. To eliminate these possibilities, we now follow [11].

To construct the prime $\mathfrak{q}_1$, we consider the following situation. Suppose that $\overline{\rho}\colon \mathrm{Gal}(\overline{F}/F) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is our given representation, and that

$$\begin{aligned} \chi : \mathrm{Gal}(\overline{F}/F) &\longrightarrow \mathbf{F}_\ell^\times \\ \mathsf{Frob}_{\mathfrak{q}} &\mapsto N_{F/\mathbf{Q}}(\mathfrak{q}) \pmod{\ell} \qquad (\text{for } \mathfrak{q} \nmid \ell) \end{aligned}$$

is the reduction of the cyclotomic character. Suppose G is a finite group through which both $\overline{\rho}$ and χ factor. Say $g \in G$ is *special* if one has $\operatorname{tr}\overline{\rho}(g)^2/\det\overline{\rho}(g) = (1+\chi(g))^2/\chi(g)$, then

LEMMA 12.3. *Suppose* $[F(\mu_\ell) : F] \geqslant 4$. *If all* $g \in G - \ker\chi$ *are special, then* $\overline{\rho}$ *is reducible.*

Proof. As in [11], Lemma 11, let $H = \overline{\rho}(G)$, and let Z be the scalar matrices in H. One easily sees that if $\overline{\rho}(g) \in Z$, then $\chi(g) = 1$. So χ factors through

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

$G/\overline{\rho}^{-1}(Z) \cong H/Z$. The kernel of χ is $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(\mu_\ell))$; an element $\mathrm{Frob}_\mathfrak{q}$ is in the kernel if and only if $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{q}) \equiv 1 \pmod{\ell}$, i.e., if ζ_ℓ is an ℓth root of unity, then $\zeta_\ell^{N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{q})} = \zeta_\ell$, which is equivalent to insisting that $\mathrm{Frob}_\mathfrak{q}$ fixes not only F, but also ℓth roots of unity. Under the given hypothesis on ℓ, one sees therefore that $|\mathrm{im}\,\chi| > 3$. However, H/Z is also a finite subgroup of $\mathrm{PGL}_2(\overline{\mathbf{F}}_\ell)$, and one has a complete classification of these subgroups. The result now follows exactly as in [11]; subgroups either have a maximal cyclic quotient of size 3 or less (a contradiction to $|\mathrm{im}\,\chi| > 3$) or give rise to reducible representations.

We note also that if $[\mathrm{F}(\mu_\ell) \colon \mathrm{F}] = 2$ (this degree must be even, so this is the only other possibility), the refinement of Diamond and Taylor [12], Lemma 3, continues to hold; that is, the conclusion of the Lemma 12.3 is satisfied if $\overline{\rho}$ is not induced from a character of $\ker\chi$.

To conclude, given that $\overline{\rho}$ is irreducible, one chooses $g \in G$ which is not special and for which $\chi(g) \neq 1$. By Čebotarev's theorem, there exist infinitely many primes $\mathfrak{q}$ such that $\mathrm{Frob}_\mathfrak{q}$ maps to g. Pick such a prime $\mathfrak{q}_1$ satisfying $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{q}_1) > 4^d$, and $\mathfrak{q}_1 \nmid \mathfrak{n}\ell\mathfrak{p}$.

It follows that Theorem 11.3 implies Mazur's Principle.

13. Sheaves

Before embarking upon the proof of Theorem 11.3, we introduce some notation following [4], Sections 2,4, in defining certain sheaves on our Shimura curves. We will always assume H sufficiently small that $\mathbf{M}_{0,H}$ exists.

Fix integers $k_1 \equiv \cdots \equiv k_d \pmod 2$, where we now insist that all $k_i \geqslant 2$. Define w_i and v_i as in [15].

Let $\mathrm{E} \subset \mathbf{C}$ be a Galois number field of finite degree containing F and splitting B.

For each $i = 1, \ldots, d$, $B \otimes_{\mathrm{F},\tau_i} \mathrm{E} \cong M_2(\mathrm{E})$, and this defines an equivalence class, written ξ_i, of representations of $B^\times = G(\mathbf{Q})$ on $W_i = \mathrm{E}^2$. Consider

$$\xi = \bigotimes_{i=1}^{d} [(\tau_i \circ \nu)^{v_i} \mathrm{Sym}^{k_i-2}(\xi_i)]$$

of $G(\mathbf{Q})$ acting on the space $W = \bigotimes_{i=1}^d W_i$. This action naturally extends to a unique action of the algebraic group G.

Write W_λ for $W \otimes_\mathrm{E} \mathrm{E}_\lambda$.

We first define a sheaf on $M_K(\mathbf{C})$, assuming K sufficiently small so that this definition makes sense, by

$$\mathcal{F}^K_{k,w}(\mathbf{Q}_\ell)_\mathbf{C} = G(\mathbf{Q}) \Big\backslash \Big(G(\mathbf{A}^\infty)/K \times X \times W_\lambda \Big)$$

with $G(\mathbf{Q})$ acting on W_λ via $G(\mathbf{Q}_\ell)$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

On $M(\mathbf{C})$, this sheaf is constant, with fibre W_λ. Thus we see that

$$\mathcal{F}^K_{k,w}(\mathbf{Q}_\ell)_\mathbf{C} = \Big(M(\mathbf{C}) \times W_\lambda\Big) \Big/ \Big(K/(K \cap \widehat{Z(\mathbf{Q})})\Big) ,$$

as $K/(K \cap \widehat{Z(\mathbf{Q})})$ is the Galois group of the covering $M(\mathbf{C}) \longrightarrow M_K(\mathbf{C})$. Here $\widehat{Z(\mathbf{Q})}$ is the closure of $Z(\mathbf{Q})$ in $Z(\mathbf{A}^\infty)$.

Choose a lattice L_λ inside W_λ. If K is sufficiently small, it stabilises L_λ, where $k \in K$ acts on the right on W_λ by $\xi(k_\ell)^{-1}$, and we pick a normal subgroup $K' \subset K$ such that K' acts trivially on $L_\lambda/\ell^n L_\lambda$.

Then we may define a sheaf $\mathcal{F}_{k,w}(\mathbf{Z}_\ell)_\mathbf{C}$ by

$$\mathcal{F}^K_{k,w}(\mathbf{Z}_\ell)_\mathbf{C} = \Big(M(\mathbf{C}) \times L_\lambda\Big) \Big/ \Big(K/(K \cap \widehat{Z(\mathbf{Q})})\Big),$$

and

$$\mathcal{F}^K_{k,w}(\mathbf{Z}/\ell^n\mathbf{Z})_\mathbf{C} = \Big(M_{K'}(\mathbf{C}) \times (L_\lambda/\ell^n L_\lambda)\Big)/(K/K').$$

Carayol then defines an étale sheaf of $(\mathbf{Z}/\ell^n\mathbf{Z})$-modules $\mathcal{F}^K_{k,w}(\mathbf{Z}/\ell^n\mathbf{Z})$ on M_K by

$$\mathcal{F}^K_{k,w}(\mathbf{Z}/\ell^n\mathbf{Z}) = \Big(M_{K'} \times (L_\lambda/\ell^n L_\lambda)\Big)/(K/K').$$

Write $\mathcal{F}^K_{k,w}(\mathbf{Z}_\ell)$ for the inverse limit of these sheaves (as n varies).

For $K = \mathrm{GL}_2(\mathcal{O}_\mathfrak{p})H$, Carayol defines, in the same way, a sheaf of $(\mathbf{Z}/\ell^n\mathbf{Z})$-modules on $\mathbf{M}_{0,H}$ by

$$\boldsymbol{\mathcal{F}}^{0,H}_{k,w}(\mathbf{Z}/\ell^n\mathbf{Z}) = \Big(\mathbf{M}_{0,H'} \times (L_\lambda/\ell^n L_\lambda)\Big)/(H/H').$$

These sheaves $\boldsymbol{\mathcal{F}}^{0,H}_{k,w}(\mathbf{Z}/\ell^n\mathbf{Z})$ are lisse étale sheaves.

Henceforth, write Λ for $\mathbf{Z}/\ell\mathbf{Z}$. Write $\boldsymbol{\mathcal{F}}$ for the sheaf $\boldsymbol{\mathcal{F}}^{0,H}_{k,w}(\Lambda)$ on $\mathbf{M}_{0,H}$. For X a scheme over $\mathbf{M}_{0,H}$ (for instance, $\mathbf{M}_{U_0(\mathfrak{p}),H}$), we denote the pull-back of $\boldsymbol{\mathcal{F}}$ to X also by $\boldsymbol{\mathcal{F}}$ (as for constant sheaves).

14. Exact Sequences

We follow Carayol [4] in defining certain exact sequences.

The exact sequence of vanishing cycles for the (proper) morphism

$$\begin{array}{c}\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \mathcal{O}_\mathfrak{p} \\ \big\downarrow \\ \operatorname{spec} \mathcal{O}_\mathfrak{p},\end{array}$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

and the sheaf $\mathcal{F}$ is then (notation as in [SGA7], XIII):

$$0 \longrightarrow H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F}) \longrightarrow H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F})$$

$$\longrightarrow \bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\mathcal{F})_x \longrightarrow H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F})$$

$$\xrightarrow{sp} H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F}) \longrightarrow \cdots,$$

where Σ_H denotes the set of singular points for the $\overline{\kappa}$-scheme $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}$; Σ_H consists of a finite number of non-degenerate quadratic points.

Write

$$L(H) = \ker(sp\colon H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F}) \longrightarrow H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F})).$$

Make the following abbreviations:

$$Z(H) = H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F})$$

$$M(H) = H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F})$$

$$X(H) = \bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\mathcal{F})_x = \bigoplus_{x \in \Sigma_H} ((R^1\Phi_{\overline{\eta}}\Lambda)_x \otimes \mathcal{F}_x).$$

The equality here holds by [SGA 7], XIII, 2.1.13. Write

$$\widetilde{X}(H) = \ker(X(H) \longrightarrow L(H)).$$

Thus we have an exact sequence

$$0 \longrightarrow Z(H) \longrightarrow M(H) \longrightarrow \widetilde{X}(H) \longrightarrow 0,$$

which we will refer to as *exact sequence (A)*.

We will write L, M, X, $\widetilde{X}$, and Z for the inductive limits (over H) of $L(H)$, $M(H)$, $X(H)$, $\widetilde{X}(H)$ and $Z(H)$.

Next, we construct a second exact sequence, based on the comparison between the cohomology of the special fibre, and the cohomology of its normalisation.

Recall from above that we have a map

$$\begin{array}{c} \mathbf{M}_{0,H} \otimes \overline{\kappa} \sqcup \mathbf{M}_{0,H} \otimes \overline{\kappa} \\ \downarrow r \\ \mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa} \end{array}$$

and $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}$ may be regarded as two copies of $\mathbf{M}_{0,H} \otimes \overline{\kappa}$ glued together transversally above each supersingular point of $\mathbf{M}_{0,H} \otimes \overline{\kappa}$. As r is an isomorphism away from supersingular points, there is an exact sequence of sheaves

$$0 \longrightarrow \mathcal{F} \longrightarrow r_*r^*\mathcal{F} \longrightarrow \mathcal{G} \longrightarrow 0,$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

where $\mathcal{G}$ is a skyscraper sheaf supported on Σ_H.

Then one gets an exact sequence

$$\begin{aligned}0 &\longrightarrow H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F}) \longrightarrow H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \mathcal{F}) \\ &\xrightarrow{\alpha} H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{G}) \longrightarrow H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{F}) \\ &\longrightarrow H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \mathcal{F}) \longrightarrow 0.\end{aligned}$$

Let

$$L'(H) = \mathrm{Im}(\alpha : H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \mathcal{F}) \longrightarrow H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{G})).$$

Make the following abbreviations:

$$\begin{aligned}Y(H) &= H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \mathcal{G}) \\ &= \bigoplus_{x \in \Sigma_H} \mathcal{G}_x, \\ R(H) &= H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \mathcal{F}).\end{aligned}$$

Write $\widetilde{Y}(H)$ for $Y(H)/L'(H)$.

Then we have an exact sequence

$$0 \longrightarrow \widetilde{Y}(H) \longrightarrow Z(H) \longrightarrow R(H) \longrightarrow 0,$$

which we will refer to as *exact sequence (B)*.

Again we will indicate the inductive limits of these terms by dropping reference to H.

Exact sequences (A) and (B) combine to give the following fundamental diagram

$$\begin{array}{ccccccccc} & & 0 & & & & & & \\ & & \downarrow & & & & & & \\ & & \widetilde{Y}(H) & & & & & & \\ & & \downarrow & & & & & & \\ 0 & \to & Z(H) & \to & M(H) & \to & \widetilde{X}(H) & \to & 0 \\ & & \downarrow & & & & & & \\ & & R(H) & & & & & & \\ & & \downarrow & & & & & & \\ & & 0 & & & & & & \end{array}$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

This diagram is implicit in Langlands [18], and Ribet [23], and a version of it is in [4].

15. Group Actions

Let Γ be the restricted direct product of the groups $(B \otimes \mathrm{F}_v)^\times$, the product being taken over all finite places of F different from $\mathfrak{p}$, and let $\Gamma_1 = \{g \in \Gamma | g_\ell = 1\}$.

There is a natural action of $G(\mathbf{A}^\infty)$ on the projective system (as K varies) of the analytic curves $\{M_K(\mathbf{C})\}$, given by right multiplication. More precisely, $g \in G(\mathbf{A}^\infty)$ induces a map $\rho_g\colon M_K \xrightarrow{\sim} M_{g^{-1}Kg}$. This action extends to an action of Γ on the projective system of the integral models of the corresponding algebraic curves ([3], 6.2): $\rho_g\colon \mathbf{M}_{0,H} \xrightarrow{\sim} \mathbf{M}_{0,g^{-1}Hg}$. For similar reasons, namely the functorial definitions of the integral models with $U_0(\mathfrak{p})$-level structure, we also obtain an action of Γ on the projective system (over H) of $\{\mathbf{M}_{U_0(\mathfrak{p}),H}\}$.

For $g \in \Gamma_1$, $g_\ell = 1$, so that the action of Γ_1 fixes L_λ, and then it is clear that the actions of Γ_1 above on the projective systems $\{\mathbf{M}_{0,H}\}$ and $\{\mathbf{M}_{U_0(\mathfrak{p}),H}\}$ lift to actions on the projective systems $\{\boldsymbol{\mathcal{F}}^{0,H}|_{\mathbf{M}_{0,H}}\}$ and $\{\boldsymbol{\mathcal{F}}^{0,H}|_{\mathbf{M}_{U_0(\mathfrak{p}),H}}\}$.

We immediately obtain an action of Γ_1 on all of the systems of cohomology groups appearing in Section 14. In particular, the inductive limits of sequences (A) and (B),

$$0 \longrightarrow Z \longrightarrow M \longrightarrow \widetilde{X} \longrightarrow 0 \quad \text{and} \quad 0 \longrightarrow \widetilde{Y} \longrightarrow Z \longrightarrow R \longrightarrow 0,$$

are equivariant for the Γ_1-action.

With the aid of this action, we can define Hecke operators on all the cohomology groups, as in [11], Section 3, or [15], (7.2).

Suppose K and K' are sufficiently small open compact subgroups of $G(\mathbf{A}^\infty)$, and $g \in G(\mathbf{A}^\infty)$ with $g_\ell = 1$. There is a natural identification of sheaves on $M_{K\cap gK'g^{-1}}$: $\mathcal{F}^{gK'g^{-1}}|_{M_{K\cap gK'g^{-1}}} = \mathcal{F}^{K\cap gK'g^{-1}}$.

Then define

$$\begin{aligned}[KgK'] \;:\; & H^i(M_{K'} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}^{K'}) \xrightarrow{\text{res}} H^i(M_{K'\cap g^{-1}Kg} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}^{K'}|_{M_{K'\cap g^{-1}Kg}}) \\ & \xrightarrow{\rho_g^*} H^i(M_{gK'g^{-1}\cap K} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}^{gK'g^{-1}}|_{M_{gK'g^{-1}\cap K}}) \\ & = H^i(M_{gK'g^{-1}\cap K} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}^{K\cap gK'g^{-1}}) \\ & \xrightarrow{Tr} H^i(M_K \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}^K).\end{aligned}$$

Let $K' = K$. If $\mathfrak{q}$ is a prime of $\mathcal{O}_{\mathrm{F}}$ which splits in B and does not divide ℓ, let $\omega_{\mathfrak{q}} \in \mathbf{A}_{\mathrm{F}}^\infty$ be such that $\omega_{\mathfrak{q}}$ is 1 at every place, except for $\mathfrak{q}$, where the component is to be a uniformiser. Then write

$$T_{\mathfrak{q}} = \left[K \begin{pmatrix} 1 & 0 \\ 0 & \omega_{\mathfrak{q}} \end{pmatrix} K\right].$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

If also $K_{\mathfrak{q}} = \mathrm{GL}_2(\mathcal{O}_{\mathfrak{q}})$, define

$$S_{\mathfrak{q}} = \left[K \begin{pmatrix} \omega_{\mathfrak{q}} & 0 \\ 0 & \omega_{\mathfrak{q}} \end{pmatrix} K \right].$$

In the same way, define actions of these Hecke operators on $H^i(\mathbf{M}_{0,H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}})$ and $H^i(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}})$.

The equivariance of Γ_1 on the inductive limits of exact sequences (A) and (B) implies the equivariance of the action of the Hecke operators above on the fundamental diagram.

Of course, there is also a Galois action on the two exact sequences. By [SGA 7], XIII, 1.3.2.2, exact sequence (A) is $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$-equivariant, and the same is also true of exact sequence (B), on which $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ acts through $\mathrm{Gal}(\mathrm{F}_{\mathfrak{p}}^{\mathrm{nr}}/\mathrm{F}_{\mathfrak{p}}) \cong \mathrm{Gal}(\overline{\kappa}/\kappa)$.

Thus the fundamental diagram is equivariant for $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$, and the inductive limit is equivariant for Γ_1.

There is an isomorphism of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$-modules ([4], Section 2.3):

$$H^1(M_K \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_{k,w}^K(\mathbf{Z}_\ell)) \otimes_{\mathcal{O}_{\mathrm{E}_\lambda}} \mathbf{C} \cong \bigoplus_{\pi \in C} (\pi^\infty)^K \otimes \rho_\lambda^\pi,$$

where ρ_λ^π denotes a 2-dimensional λ-adic Galois representation associated to π. Recall that we fixed in Section 11 a modulo ℓ representation $\overline{\rho}$. As there is an embedding

$$H^1(M_K \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_{k,w}^K(\mathbf{Z}_\ell)) \otimes \mathbf{Z}/\ell\mathbf{Z} \hookrightarrow H^1(M_K \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_{k,w}^K(\mathbf{Z}/\ell\mathbf{Z})),$$

arising from the short exact sequence

$$0 \longrightarrow \mathcal{F}_{k,w}^K(\mathbf{Z}_\ell) \xrightarrow{\ell} \mathcal{F}_{k,w}^K(\mathbf{Z}_\ell) \longrightarrow \mathcal{F}_{k,w}^K(\mathbf{Z}/\ell\mathbf{Z}) \longrightarrow 0,$$

it follows that $\overline{\rho}$ is a submodule of $H^1(M_K \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_{k,w}^K(\mathbf{Z}/\ell\mathbf{Z})) \otimes \overline{\mathbf{F}}_\ell$, and then $\overline{\rho}|_{W_{\mathrm{F}_{\mathfrak{p}}}}$ will be a submodule of $H^1(M_K \otimes_{\mathrm{F}_{\mathfrak{p}}} \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F}_{k,w}^K(\mathbf{Z}/\ell\mathbf{Z})) \otimes \overline{\mathbf{F}}_\ell$.

16. Analysis of the Fundamental Diagram

We now analyse $R(H)$.

LEMMA 16.1. $H^i(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \boldsymbol{\mathcal{F}}) \cong H^i(\mathbf{M}_{0,H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}})^2$.

Proof. We note that $r\colon \mathbf{M}_{0,H} \otimes \overline{\kappa} \sqcup \mathbf{M}_{0,H} \otimes \overline{\kappa} \longrightarrow \mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}$ is a finite morphism (from Theorem 10.2, each component is finite and flat over $\mathbf{M}_{0,H} \otimes \overline{\kappa}$).

This isomorphism is Γ_1-equivariant.

Remark. 16.2. We also have an isomorphism

$$H^i(\mathbf{M}_{0,H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \cong H^i(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}}).$$

This is a well-known result and follows from the fact that $\mathbf{M}_{0,H}$ has good reduction. It may also be seen as a degenerate case of the vanishing cycle sequence, as $\mathbf{M}_{0,H} \otimes \overline{\kappa}$ has no singularities. It is thus equivariant for $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ and for Γ_1.

COROLLARY 16.3. $R(H) \cong H^1(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}})^2$.

17. Monodromy

The action of the inertia group $I = I_{\mathrm{F}_{\mathfrak{p}}}$ on $M(H) = H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}})$ is computed by means of the following diagram [SGA 7], XIII, 2.4.6:

$$\begin{array}{ccc} H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}}) & \xrightarrow{\sigma - 1} & H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}}) \\ \downarrow & & \uparrow \\ \bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\boldsymbol{\mathcal{F}})_x & \xrightarrow{\bigoplus \mathrm{Var}_x(\sigma)} & \bigoplus_{x \in \Sigma_H} \mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}). \end{array}$$

Recall that the terms at the top of this diagram are $M(H)$, and that the bottom left-hand corner of this diagram is none other than $X(H)$. Further, the bottom map is given by

$$\bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\Lambda)_x \otimes \boldsymbol{\mathcal{F}}_x \xrightarrow{\bigoplus \mathrm{Var}_x(\sigma) \otimes 1} \bigoplus_{x \in \Sigma_H} \mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\Lambda) \otimes \boldsymbol{\mathcal{F}}_x.$$

$\mathrm{Var}_x(\sigma)$ is described by [SGA 7], XV, 3.3.5 and 3.4(iii), the Picard–Lefschetz formula.

Given $x \in \Sigma_H$, we define a character ϵ_x as follows. By [SGA 7], XV, 1.3.1(i), the complete local ring of $\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \widehat{\mathcal{O}}^{\mathrm{nr}}_{\mathfrak{p}}$ at x is isomorphic to $\widehat{\mathcal{O}}^{\mathrm{nr}}_{\mathfrak{p}}[[X,Y]]/(Q(X,Y) - b)$, where $Q(X,Y) = XY +$ higher order terms.

We define

$$\begin{aligned} \epsilon_x \colon I &\longrightarrow \Lambda(1) \\ \sigma &\mapsto \epsilon_x(\sigma) = \sigma(b^{1/\ell})/b^{1/\ell}. \end{aligned}$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

By ([SGA 7], XV, 3.3.3), $\mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\Lambda)$ has a natural generator $\pm\delta_x$, well-defined up to sign, and for $\sigma \in I$, and $a \in (R^1\Phi_{\overline{\eta}}\Lambda)_x$, one has

$$\mathrm{Var}_x(\sigma)(a) = -\epsilon_x(\sigma)(a, \delta_x)\delta_x,$$

where $(\ ,\ )$ is a perfect pairing into $\Lambda(-1)$ ([SGA 7], XV, 2.2.5(C)). Write N_x for the morphism $N_x(a) = (a, \delta_x)\delta_x$. As $(\ ,\ )$ is a perfect pairing, N_x gives an isomorphism between the spaces $(R^1\Phi_{\overline{\eta}}\Lambda(1))_x$ and $\mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\Lambda)$, as $(\ , \delta_x)$ is surjective onto $\Lambda(-1)$.

Finally, as $\mathbf{M}_{U_0(\mathfrak{p}),H}$ is regular, it follows that at each $x \in \Sigma_H$, one has $v_{\mathfrak{p}}(b) = 1$. Then the character $\epsilon_x(\sigma)$ is independent of the particular singularity chosen.

Write N for the induced morphism:

$$\begin{array}{ccc}
H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}})(1) & \xrightarrow{\quad N \quad} & H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}}) \\
\downarrow & & \uparrow \\
\bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\boldsymbol{\mathcal{F}}(1))_x & \xrightarrow{\bigoplus N_x} & \bigoplus_{x \in \Sigma_H} \mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}).
\end{array}$$

Then $\sigma \in I_{\mathrm{F}_{\mathfrak{p}}}$ acts by $1 - \epsilon(\sigma)N$.

Next, we interpret the image of the variation map.

We consider exact sequences (A) and (B) in cohomology with support in Σ_H. Σ_H is fixed by the action of $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ ([3], 10.3), and $g \in \Gamma_1$ maps Σ_H to $\Sigma_{g^{-1}Hg}$ ([4]), so that as the exact sequences of sheaves that give rise to these long exact cohomology sequences are equivariant for the actions of these groups, it follows that the sequences with support in Σ_H are also equivariant for the Galois action, and the inductive systems of these sequences are equivariant for the action of Γ_1.

LEMMA 17.1. *There is an isomorphism*

$$H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \xrightarrow{\sim} \bigoplus_{x \in \Sigma_H} \mathbf{H}^1_{\{x\}}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}).$$

Proof. We consider exact sequence (A) with support in Σ_H.

$$\begin{aligned}
0 &\longrightarrow H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \\
&\longrightarrow \mathbf{H}^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}) \xrightarrow{\phi} \bigoplus_{x \in \Sigma_H} (R^1\Phi_{\overline{\eta}}\boldsymbol{\mathcal{F}})_x \longrightarrow \cdots
\end{aligned}$$

However, the map ϕ is the zero map ([SGA 7], XV, (2.2.5.8) or 3.4). Thus there is an isomorphism

$$H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \xrightarrow{\sim} \mathbf{H}^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}),$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

as required.

We now prove a lemma which is presumably well known, but for which we could not find a reference.

LEMMA 17.2. *Let $X_1 \xrightarrow{r} X_2$ be a finite morphism of schemes. Consider the following Cartesian diagram:*

$$\begin{array}{ccc} Z_1 & \xrightarrow{s} & Z_2 \\ \downarrow{\scriptstyle i_1} & & \downarrow{\scriptstyle i_2} \\ X_1 & \xrightarrow{r} & X_2 \end{array}$$

in which i_1 and i_2 are closed immersions. Let $\mathcal{F}$ be an étale sheaf on X_1. Then $H^i_{Z_1}(X_1, \mathcal{F}) \cong H^i_{Z_2}(X_2, r_*\mathcal{F})$.

Proof. As r is finite, r_* is exact ([20], II3.6). Further, r_* takes injectives to injectives ([20], III1.2). Thus it suffices to prove the result for $i = 0$. We have

$$H^0_{Z_1}(X_1, \mathcal{F}) = H^0(X_1, i_{1*}i_1^!\mathcal{F})$$

(by [20], p.91)

$$= H^0(X_2, r_*i_{1*}i_1^!\mathcal{F})$$

(as r is finite)

$$= H^0(X_2, i_{2*}s_*i_1^!\mathcal{F})$$

$$\cong H^0(X_2, i_{2*}i_2^!r_*\mathcal{F})$$

(by [SGA 4], XVIII, 3.1.12.3, as closed immersions are compactifiable [SGA 4], XVII, 3.2.3(i))

$$\cong H^0_{Z_2}(X_2, r_*\mathcal{F}),$$

as required.

Exact sequence (B), with cohomology in Σ_H, becomes:

$$\cdots \longrightarrow H^0_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_*r^*\boldsymbol{\mathcal{F}}) \longrightarrow H^0_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{G}})$$
$$\longrightarrow H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \longrightarrow H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_*r^*\boldsymbol{\mathcal{F}}) \longrightarrow \cdots$$

LEMMA 17.3. *There is an isomorphism*

$$H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{G}}) \cong H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}).$$

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

Proof. The map $r\colon \mathbf{M}_{0,H}\otimes\overline{\kappa}\sqcup\mathbf{M}_{0,H}\otimes\overline{\kappa}\longrightarrow\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa}$ is finite, so, by Lemma 17.2, there is an isomorphism:

$$H^i_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa}, r_*r^*\boldsymbol{\mathcal{F}})\cong H^i_{\Sigma^0_H}(\mathbf{M}_{0,H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}})^2,$$

where Σ^0_H denotes the set of supersingular points of $\mathbf{M}_{0,H}\otimes\overline{\kappa}$.

Further, $H^i_{\Sigma^0_H}(\mathbf{M}_{0,H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}})=(0)$ for $i=0$ and $i=1$, as $\mathbf{M}_{0,H}\otimes\overline{\kappa}$ is smooth, and $\boldsymbol{\mathcal{F}}$ is locally constant ([20], VI 5.1).

We conclude that there is an isomorphism

$$H^0_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{G}})\cong H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}}),$$

and as $\boldsymbol{\mathcal{G}}$ is supported on Σ_H, we finally deduce the desired isomorphism.

For purely functorial reasons, we have commutative squares:

$$\begin{array}{ccc} H^0_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{G}}) & \xrightarrow{\sim} & H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}}) \\ \Vert & & \downarrow \\ Y(H)=H^0(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{G}}) & \longrightarrow & H^1(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}})=Z(H) \end{array}$$

coming from exact sequence (B), and

$$\begin{array}{ccc} H^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}}) & \xrightarrow{\sim} & \mathbf{H}^1_{\Sigma_H}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}}) \\ \downarrow & & \downarrow \\ Z(H)=H^1(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},\boldsymbol{\mathcal{F}}) & \longrightarrow & \mathbf{H}^1(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}})=M(H) \end{array}$$

coming from exact sequence (A). It follows that the map

$$Y(H)\longrightarrow Z(H)\longrightarrow M(H)$$

coming from exact sequences (A) and (B) coincides with the map

$$Y(H)\xrightarrow{\sim}\bigoplus_{x\in\Sigma_H}\mathbf{H}^1_{(x)}(\mathbf{M}_{U_0(\mathfrak{p}),H}\otimes\overline{\kappa},R\Psi_{\overline{\eta}}\boldsymbol{\mathcal{F}})\longrightarrow M(H)$$

coming from the monodromy theory.

Combining the isomorphisms (17.1) and (17.3), we deduce the main result of the section:

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

PROPOSITION 17.4. *We have the following diagrams:*

$$\begin{array}{ccc} M(H) & \xrightarrow{\sigma-1} & M(H) \\ \big\downarrow & & \big\uparrow \\ X(H) & \xrightarrow{\mathrm{Var}(\sigma)} & Y(H) \end{array}$$

for $\sigma \in I$*, and*

$$\begin{array}{ccc} M(H)(1) & \xrightarrow{N} & M(H) \\ \big\downarrow & & \big\uparrow \\ X(H)(1) & \xrightarrow{\bigoplus N_x} & Y(H). \end{array}$$

18. Proof of Theorem 11.3

We begin by recalling the statement of Theorem 11.3, whose proof is the object of this section.

THEOREM 11.3. *Assume* $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$*, a continuous irreducible semisimple representation, is attached to a Hilbert cuspidal eigenform* $f \in S_{k,w}(U_0(\mathfrak{p}) \cap U)$*, where* $\mathfrak{p} \nmid \ell$*,* $k \geqslant 2t$*, and* U *is sufficiently small. Suppose* U *decomposes as* $\prod_{\mathfrak{q}} U_{\mathfrak{q}}$*, with* $U_{\mathfrak{p}} = \mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}})$*. If* $[\mathrm{F} : \mathbf{Q}]$ *is even, suppose also that there exists some finite place* $\mathfrak{q}_0 \neq \mathfrak{p}$ *of* F *at which the automorphic representation corresponding to* f *is special or supercuspidal. Then if* $\overline{\rho}$ *is irreducible, and unramified at* $\mathfrak{p}$*, and* $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$*, there exists a Hilbert cuspidal eigenform* $f' \in S_{k,w}(U)$ *to which* $\overline{\rho}$ *is attached.*

In Section 11, we fixed some $\overline{\rho}\colon \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$, satisfying the conditions of the theorem.

Let $\mathbf{T}$ denote the Hecke algebra generated by the Hecke operators $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ for all $\mathfrak{q} \nmid \mathfrak{p}\ell$ such that $U_{\mathfrak{q}} = \mathrm{GL}_2(\mathcal{O}_{\mathfrak{q}})$. From Section 15, $\mathbf{T}$ acts on the fundamental diagram. Let $\mathfrak{m}$ denote the maximal ideal of $\mathbf{T}$ corresponding to f. We will write $\mathbf{F}$ for $\mathbf{T}/\mathfrak{m}$, so that $\overline{\rho}$ has a model valued in $\mathrm{GL}_2(\mathbf{F})$.

We fixed in Section 11 a quaternion algebra B, such that, under the assumptions of the theorem, the Jacquet–Langlands correspondence provides an automorphic representation π on B associated to (the automorphic representation corresponding to) f. We also fixed, for every place v at which B is split, an isomorphism $B \otimes_{\mathrm{F}} \mathrm{F}_v \cong M_2(\mathrm{F}_v)$.

Under the assumption that U is sufficiently small, π will have fixed vectors under $U_0(\mathfrak{p}).H$ for some open compact subgroup $H \subset \Gamma$ corresponding to U, sufficiently small in the sense that an integral model for $M_{0,H}$ (and thus $M_{U_0(\mathfrak{p}),H}$) exists.

From Section 15, $\overline{\rho}|_{W_{\mathrm{F}_\mathfrak{p}}}$ is a submodule of

$$M(H) \otimes_{\mathbf{F}_\ell} \overline{\mathbf{F}}_\ell = H^1(M_{U_0(\mathfrak{p}),H} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \boldsymbol{\mathcal{F}}) \otimes_{\mathbf{F}_\ell} \overline{\mathbf{F}}_\ell.$$

Recall that we have the fundamental diagram

$$\begin{array}{ccccccccc} & & 0 & & & & & & \\ & & \downarrow & & & & & & \\ & & \widetilde{Y}(H) & & & & & & \\ & & \downarrow & & & & & & \\ 0 & \to & Z(H) & \to & M(H) & \to & \widetilde{X}(H) & \to & 0 \\ & & \downarrow & & & & & & \\ & & R(H) & & & & & & \\ & & \downarrow & & & & & & \\ & & 0 & & & & & & \end{array}$$

together with an isomorphism (17.4) $N\colon X(H)(1) \xrightarrow{\sim} Y(H)$. Everything is equivariant for the action of $\mathrm{Gal}(\overline{\mathrm{F}}_\mathfrak{p}/\mathrm{F}_\mathfrak{p})$, Γ_1 and $\mathbf{T}$.

We will consider this diagram as a diagram of $\mathbf{T}$-modules, and localise at the maximal ideal $\mathfrak{m}$.

LEMMA 18.1. $L'(H)_\mathfrak{m} = 0$.

Proof. One has an isomorphism

$$H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \boldsymbol{\mathcal{F}}) \cong H^0(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \boldsymbol{\mathcal{F}})^2.$$

One knows ([3], Section 2) that

$$\pi_0(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_\mathfrak{p}) \cong \mathrm{F}_+^\times \backslash (\mathbf{A}_\mathrm{F}^\infty)^\times / (\nu(H) \times \mathcal{O}_\mathfrak{p}^\times).$$

Then

$$\pi_0(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_\mathfrak{p}) \cong (\mathrm{F}^\mathfrak{p})_+^\times \backslash (\mathbf{A}_\mathrm{F}^{\infty,\mathfrak{p}})^\times / \nu(H),$$

where we write $\mathrm{F}^\mathfrak{p}$ for $\mathrm{F} \cap \mathcal{O}_\mathfrak{p}$, and $G(\mathbf{A}^{\infty,p,\ell}) \subset \Gamma_1$ acts on the set of components through $(\mathbf{A}_\mathrm{F}^{\infty,p,\ell})^\times$ via ν (this is [3], 1.3).

But then maximal ideals in $\mathbf{T}$ lying in the support of $L'(H)$ must correspond to 1-dimensional automorphic representations, as cuspidal representations on quaternionic groups admit infinite-dimensional components at almost every place, and thus do not factor through the norm ([4], 4.4).

LEMMA 18.2. $L(H)_{\mathfrak{m}} = 0$.

Proof. There is an isomorphism

$$H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, \boldsymbol{\mathcal{F}}) \cong H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \boldsymbol{\mathcal{F}}),$$

as $\boldsymbol{\mathcal{G}}$ is concentrated at points, so that its cohomology groups vanish in degree $\geqslant 1$. By 16.2, we may regard this latter group as $H^2(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}})^2$. This is Poincaré dual to the group $H^0(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \check{\boldsymbol{\mathcal{F}}}(1))^2$. But the analysis of the previous lemma ensures that the action of $G(\mathbf{A}^{\infty,p,\ell})$ factors through $(\mathbf{A}_{\mathrm{F}}^{\infty,p,\ell})^\times$. The result follows as before.

Assumption. We assume for a contradiction that there is no automorphic representation π' of G with a fixed vector under $\mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}}).H$ which gives rise to $\overline{\rho}$.

LEMMA 18.3. *Under this assumption*, $R(H)_{\mathfrak{m}} = (0)$.

Proof. Recall that $R(H) \cong H^1(\mathbf{M}_{0,H} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \boldsymbol{\mathcal{F}})^2$. But this implies that all cuspidal automorphic representations π contributing to $R(H)$ satisfy $(\pi^\infty)^{\mathrm{GL}_2(\mathcal{O}_{\mathfrak{p}}).H} \neq (0)$. Under the assumption, no such automorphic representation gives rise to $\overline{\rho}$. It follows that $R(H)_{\mathfrak{m}} = (0)$. For this, note that we have a short exact sequence of sheaves

$$0 \longrightarrow r_* r^* \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell) \xrightarrow{\ell} r_* r^* \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell) \longrightarrow r_* r^* \boldsymbol{\mathcal{F}} \longrightarrow 0,$$

and part of the associated long exact sequence is:

$$H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell))_{\mathfrak{m}} \longrightarrow R(H)_{\mathfrak{m}}$$
$$\longrightarrow H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes \overline{\kappa}, r_* r^* \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell))_{\mathfrak{m}},$$

in which the first and last terms vanish. The first vanishes by the assumption and the second by a similiar argument to that of Lemma 18.2.

Then $\widetilde{X}(H)_{\mathfrak{m}} = X(H)_{\mathfrak{m}}$, and $\widetilde{Y}(H)_{\mathfrak{m}} = Y(H)_{\mathfrak{m}}$.
It follows that $Y(H)_{\mathfrak{m}} \cong Z(H)_{\mathfrak{m}}$, and we identify these spaces.
Then the fundamental diagram, after localising, becomes:

$$0 \longrightarrow Y(H)_{\mathfrak{m}} \longrightarrow M(H)_{\mathfrak{m}} \longrightarrow X(H)_{\mathfrak{m}} \longrightarrow 0,$$

together with an isomorphism $N\colon (X(H)(1))_{\mathfrak{m}} \xrightarrow{\sim} Y(H)_{\mathfrak{m}}$.

Let $\mathbf{F} = \mathbf{T}/\mathfrak{m}$, and tensor the above exact sequence with $\mathbf{F}$ to obtain:

$$Y(H) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\alpha} M(H) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\beta} X(H) \otimes_{\mathbf{T}} \mathbf{F} \longrightarrow 0,$$

together with an isomorphism $N\colon X(H)(1) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\sim} Y(H) \otimes_{\mathbf{T}} \mathbf{F}$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

LEMMA 18.4. *Under the assumption,* $H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}}) \otimes_{\mathbf{T}} \mathbf{F} \cong \overline{\rho}^a$ *for some* $a \in \mathbf{Z}_{\geqslant 1}$.

Proof. Write $V = H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell))_{\mathfrak{m}}$.

By [4], $\mathsf{Frob}_{\mathfrak{q}}^2 - T_{\mathfrak{q}}\mathsf{Frob}_{\mathfrak{q}} + N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{q})S_{\mathfrak{q}}$ annihilates $V \otimes \overline{\mathbf{Q}}_\ell$, at least for all $\mathfrak{q}$ outside some finite set of primes. It follows that the same holds for the group $V \otimes \mathbf{Q}_\ell$. From the exact sequence

$$0 \longrightarrow \mathbf{Z}_\ell \longrightarrow \mathbf{Q}_\ell \longrightarrow \mathbf{Q}_\ell/\mathbf{Z}_\ell \longrightarrow 0,$$

we deduce that we have an exact sequence containing:

$$\cdots \longrightarrow H^0(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}}(\mathbf{Q}_\ell/\mathbf{Z}_\ell))_{\mathfrak{m}} \longrightarrow V \longrightarrow V \otimes \mathbf{Q}_\ell \longrightarrow \cdots.$$

However, the first term vanishes by a similar argument to that of Lemma 18.1. Thus V is also annihilated.

Finally, one uses the sequence $0 \longrightarrow \mathbf{Z}_\ell \xrightarrow{\ell} \mathbf{Z}_\ell \longrightarrow \mathbf{Z}/\ell\mathbf{Z} \longrightarrow 0$ to deduce that $V \otimes_{\mathbf{Z}_\ell} \mathbf{Z}/\ell\mathbf{Z}$ differs from $H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}})_{\mathfrak{m}}$ only by the torsion in $H^2(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}}(\mathbf{Z}_\ell))_{\mathfrak{m}}$, and this group vanishes as in Lemma 18.2.

Thus one has an Eichler–Shimura relationship on $H^1(\mathbf{M}_{U_0(\mathfrak{p}),H} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \boldsymbol{\mathcal{F}}) \otimes_{\mathbf{T}} \mathbf{F}$, and the main theorem of [2] completes the proof.

COROLLARY 18.5. *Under the assumption,* $\overline{\rho}$ *is unramified at* $\mathfrak{p}$ *if and only if* $M(H) \otimes_{\mathbf{T}} \mathbf{F}$ *is unramified at* $\mathfrak{p}$.

We are, however, assuming in the hypotheses of Theorem 11.3 that $\overline{\rho}$ is unramified at $\mathfrak{p}$.

LEMMA 18.6. *Under the assumption, there is an isomorphism*

$$M(H) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\sim} X(H) \otimes_{\mathbf{T}} \mathbf{F}.$$

Proof. We can calculate the action of the inertia group $I = I_{\mathrm{F}_{\mathfrak{p}}}$; $\sigma \in I$ acts by

$$\begin{array}{ccc} M(H) \otimes_{\mathbf{T}} \mathbf{F} & \xrightarrow{\sigma-1} & M(H) \otimes_{\mathbf{T}} \mathbf{F} \\ \downarrow \beta & & \uparrow \alpha \\ X(H) \otimes_{\mathbf{T}} \mathbf{F} & \xrightarrow{\mathrm{Var}(\sigma)} & Y(H) \otimes_{\mathbf{T}} \mathbf{F}. \end{array}$$

By Corollary 18.5, $M(H) \otimes_{\mathbf{T}} \mathbf{F}$ is unramified. Thus $\alpha\ \mathrm{Var}(\sigma)\beta = 0$ for all $\sigma \in I$. But $\mathrm{Var}(\sigma) = -\epsilon(\sigma)N$. It follows that $\alpha N \beta = 0$.

But β is a surjection, and N is an isomorphism. It follows that $\alpha = 0$. Thus β is an isomorphism as required.

LEMMA 18.7. *Under the assumption,* $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ *acts by scalars on* $M(H) \otimes_{\mathbf{T}} \mathbf{F}$.

https://doi.org/10.1023/A:1000600311268 Published online by Cambridge University Press

Proof. One now has an isomorphism $M(H) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\sim} X(H) \otimes_{\mathbf{T}} \mathbf{F}$, and so it suffices to show that $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ acts by a scalar on $X(H) \otimes_{\mathbf{T}} \mathbf{F}$. As one also has the isomorphism $N\colon X(H)(1) \otimes_{\mathbf{T}} \mathbf{F} \xrightarrow{\sim} Y(H) \otimes_{\mathbf{T}} \mathbf{F}$, we need only show that $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ acts by a scalar on $Y(H) \otimes_{\mathbf{T}} \mathbf{F}$. But the action of $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ on $Y(H)_{\mathfrak{m}}$ may be computed in the same way as [4], Section 6, and one finds that $\mathrm{Frob}_{\mathfrak{p}}$ acts by a scalar. The result follows.

COROLLARY 18.8. *Under the assumption,* $\mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ *acts by scalars on* $\overline{\rho}$.

However, as π is special unramified at $\mathfrak{p}$, we know by [4], Théorème (A), that

$$\overline{\rho}(\sigma) \sim \begin{pmatrix} \zeta N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) & * \\ 0 & \zeta \end{pmatrix}$$

for some ζ and where $\sigma \in \mathrm{Gal}(\overline{\mathrm{F}}_{\mathfrak{p}}/\mathrm{F}_{\mathfrak{p}})$ lies above $\mathrm{Frob}_{\mathfrak{p}}$. This cannot be a scalar unless we have the condition $N_{\mathrm{F}/\mathbf{Q}}(\mathfrak{p}) \equiv 1 \pmod{\ell}$.

Thus we obtain a contradiction, and so the assumption is false. This concludes the proof of Theorem 11.3, and thus of Mazur's Principle.

Acknowledgements

It is a great pleasure to thank Kevin Buzzard, Henri Carayol, Fred Diamond, Stefan Kühnlein and especially Richard Taylor for their helpful comments. In addition, we warmly thank the referee for suggesting numerous improvements to an earlier manuscript.

References

1. Altman, A. and Kleiman, S.: *Introduction to Grothendieck Duality Theory*, Lecture Notes in Math. 146, Springer, New York, 1970.
2. Boston, N., Lenstra jr., H. and Ribet, K.: Quotients of group rings arising from two-dimensional representations, *C.R. Acad. Sci. Paris* **312** (1991), 323–328.
3. Carayol, H.: Sur la mauvaise réduction des courbes de Shimura, *Compositio Math.* **59** (1986), 151–230.
4. Carayol, H.: Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert, *Ann. Sci. Ecole. Norm. Sup. (4)* **19** (1986), 409–468.
5. Carayol, H.: Sur les représentations galoisiennes modulo ℓ attachées aux formes modulaires, *Duke Math. J.* **59** (1989), 785–801.
6. Chevalley, C.: Deux théorèmes d'arithmétique, *J. Math. Soc. Japan* **3** (1951), 36–44.
7. Darmon, H., Diamond, F. and Taylor, R.: Fermat's last theorem, in *Current Developments in Mathematics*, International Press, 1995, pp. 1–154.
8. Deligne, P.: Formes modulaires et représentations ℓ-adiques, *Sém. Bourbaki* 355 *(1968/69)*, Lecture Notes in Math. 179, Springer, New York, 1971, pp. 139–172.
9. Deligne, P. and Rapoport, M.: Les schémas de modules de courbes elliptiques, in: *Modular Functions of One Variable II*, Lecture Notes in Math. 349, Springer, New York, 1973, pp. 143–316.
10. Diamond, F.: The refined conjecture of Serre, in: *Elliptic Curves, Modular Forms and Fermat's Last Theorem*, (1995), 22–37.

11. Diamond, F. and Taylor, R.: Non-optimal levels of mod ℓ modular representations, *Invent. Math.* **115** (1994), 435–462.
12. Diamond, F. and Taylor, R.: Lifting modular mod ℓ representations, *Duke Math. J.* **74** (1994), 253–269.
13. Drinfeld, V.: Elliptic modules, *USSR Sb.* **23** (1974), 561–592.
14. Grothendieck, A.: Techniques de construction et théorèmes d'existence en géométrie algébrique III : preschémas quotients, *Sém. Bourbaki* 212 *(1960/61)*, (1961).
15. Hida, H.: On p-adic Hecke algebras for GL_2 over totally real fields, *Ann. Math.* **128** (1988), 295–384.
16. Jarvis, F.: Rigidity of moduli problems with small level structures, *in preparation* 1997.
17. Katz, N. and Mazur, B.: *Arithmetic Moduli of Elliptic Curves*, Ann. of Math. Stud. 108, Princeton Univ. Press, 1985.
18. Langlands, R.: Modular forms and ℓ-adic representations, in: *Modular Functions of One Variable* II, Lecture Notes in Math. 349, Springer, New York, 1973, pp. 361–500.
19. Matsumura, H.: *Commutative Ring Theory*, Cambridge Univ. Press, 1986.
20. Milne, J.: *Étale Cohomology*, Princeton Univ. Press, 1980.
21. Pappas, G.: Arithmetic models for Hilbert-Blumenthal abelian varieties, Ph.D. thesis, Columbia Univ., 1992.
22. Ribet, K.: On modular representations of $Gal(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms, *Invent. Math.* **100** (1990), 431–476.
23. Ribet, K.: Report on mod ℓ representations of $Gal(\overline{\mathbf{Q}}/\mathbf{Q})$, in *Proc. Symp. Pure Math*. **55** (2), Amer. Math. Soc., Providence, 1994, pp. 639–676.
24. Serre, J.-P.: Sur les représentations modulaires de degré 2 de $Gal\ (\overline{\mathbf{Q}}/\mathbf{Q})$, *Duke J. Math.* **54** (1987), 179–230.
25. Taylor, R.: On Galois representations associated to Hilbert modular forms, *Invent. Math.* **98** (1989), 265–280.