TAIWANESE JOURNAL OF MATHEMATICS
Vol. 22, No. 2, pp. 275–300, April 2018
DOI: 10.11650/tjm/8147

# A Note on Modularity Lifting Theorems in Higher Weights

Yih-Jeng Yu

Abstract. We follow the ideas of Khare and Ramakrishna-Khare and prove the modularity lifting theorem in higher weights. This approach somehow differs from that using Taylor-Wiles systems.

## 1. Introduction

The conjectural relation between Galois representations and automorphic forms is among the most important in number theory. The primary method used to establish such a relationship rests on the work of Wiles [21] and Taylor-Wiles [20].

**Theorem 1.1** (Modularity Lifting Theorem [20, 21])**.** *Let $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O})$ be an odd, continuous, absolutely irreducible, p-adic Galois representation which is ramified at finitely many primes, and de Rham at p with Hodge-Tate weights $(k-1, 0)$ with $k \geq 2$. If the reduction modulo p of $\rho$ is modular, then $\rho$ is isomorphic to an integral model of a p-adic representation $\rho_f$ arising from a newform f.*

This is sometimes described as "$R = \widehat{\mathbb{T}}$"-theorems, where $R$ is the universal deformation ring of the reduction $\overline{\rho}$ of $\rho$ and $\widehat{\mathbb{T}}$ is a certain localized Hecke algebra. The method of Taylor-Wiles for establishing this isomorphism has been further refined by many people; see for example [9].

The focus of this article is to give a different approach to proving modularity lifting theorems of Galois representations. Specifically, we generalize the approach introduced by Khare and Ramakrishna [13, 14] from weight 2 to higher weights $k < p$. The following theorem is the main result of this paper (see Theorem 4.3).

**Theorem 1.2.** *Let $N$ be a square-free positive integer, $p > 5$ be a prime not dividing $N$, and $k$ be an integer with $2 \leq k < p$. Let $f \in S_k(\Gamma_0(N))$ be a cusp newform. Let $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be the* mod$p$ *Galois representation attached to $f$. If $\overline{\rho}$ is irreducible, minimally ramified at primes dividing $Np$, and $\overline{\rho}|_{I_p} = \begin{pmatrix} \overline{\chi}_p^{k-1} & * \\ 0 & 1 \end{pmatrix}$ with $* \neq 0$, where $\overline{\chi}_p$ is the* mod$p$ *cyclotomic character, then the universal deformation ring $R$ associated to $\overline{\rho}$ is canonically isomorphic to $\widehat{\mathbb{T}}_\emptyset$ ($\simeq \widehat{\mathbb{T}}$).*

Received February 19, 2017; Accepted June 11, 2017.
Communicated by Yi-Fan Yang.
2010 *Mathematics Subject Classification.* 11F80, 11G18, 14D05, 14F05, 14G35.
*Key words and phrases.* modularity lifting theorems, Hecke algebras, Shimura curves, vanishing cycles, monodromy, Galois representations, Ramakrishna-Khare systems, Čerednik-Drinfel'd uniformizations.

The paper is organized as follows. We first review the basic properties of deformation ring $R_Q$ for an odd continuous absolutely irreducible Galois representation $\overline{\rho}$ with values in $\mathrm{GL}_2(\mathbb{F})$ where $\mathbb{F}$ is a finite field of characteristic $p$. The ring $R_Q$ is universal for the deformations of $\overline{\rho}$ unramified outside $S \cup Q$ and minimally ramified on $S$, where $S$ is the set of primes at which $\overline{\rho}$ is ramified and $Q$ is a set of auxiliary primes. According to Khare and Ramakrishna [14], we define for any subset $\alpha \subseteq Q$ a quotient $R_Q^\alpha$ of $R_Q$ which is universal for the deformations $\rho$ of $\overline{\rho}$ such that, for $q \in \alpha$, the local representation $\rho|_{G_q}$ is special, i.e., it is of the form $\left(\begin{smallmatrix} \chi_p & * \\ 0 & 1 \end{smallmatrix}\right)$ where $\chi_p$ is the $p$-adic cyclotomic character. With the result at hand, we can identify the tangent space of $R_Q^\alpha$ with a suitable Selmer group (cf. Proposition 2.4, Remark 5.3, and also [14, Lemma 16]).

Taking up the study of the bad reduction of the Shimura curves at a prime $r$ dividing the level in question by the Tate-Oort theory, we obtain an explicit description of the special fiber as the union of exactly two irreducible components of multiplicity 1 and $r-1$ respectively. These two components cross transversally at the supersingular points and nowhere else. This description enables us calculate the vanishing cycles (Proposition 3.5) and the cohomology of Shimura curves (Proposition 3.7).

Suppose $\overline{\rho}$ is modular with weight $k < p$. The first step (Proposition 4.2) is to show that there exists a set $Q$ of auxiliary primes such that $R_Q^D$ is isomorphic to $W(\mathbb{F})$ and that the corresponding deformation $\rho_Q^D$ is modular. The existence of $Q$ is due to Khare and Ramakrishna [14]. The proof of this isomorphism and the modularity of $\rho_Q^D$ use the fact that the tangent space of $R_Q^D$ is trivial and the work of Diamond-Taylor [10].

The second step (Theorem 4.3) is to show that the deformation $\rho_Q$ parametrized by $R_Q$ is modular. As in the founder article of Wiles, we introduce a localized Hecke algebra $\widehat{\mathbb{T}}_Q$ parametrizing a modular deformation of geometric origin and show that the canonical homomorphism $R_Q \to \widehat{\mathbb{T}}_Q$ is actual an isomorphism. In the proof of Theorem 4.3, we use a variant of Wiles' numerical criterion refined by Lenstra (Theorem 5.1). Thus, we are quickly reduced to study how a certain congruence module grows as one relaxes conditions of newness at primes in $Q$. Let $\pi\colon \widehat{\mathbb{T}}_Q \to \widehat{\mathbb{T}}_Q^Q \simeq R_Q^Q \simeq W(\mathbb{F})$ be the canonical homomorphism resulting from the first step, $\phi\colon R_Q \to \widehat{\mathbb{T}}_Q$, $\Phi = \ker(\pi\phi)/\ker(\pi\phi)^2$ and $\eta = \pi(\mathrm{Ann}_{\mathbb{T}}(\ker(\pi)))$. The criterion consists with verifying the equality $|W(\mathbb{F})/\eta| = |\Phi|$.

The verification of this equality is the main part of this paper. First, we need to calculate the Galois cohomology and identify $\Phi$ with certain Selmer groups; we thus obtain an upper bound for $|\Phi|$ (Proposition 5.2). Then in the proof of theorem in weight 2, Khare used a result of Ribet-Takahashi [19] which generalized a calculation of Ribet [18] in his work on Serre's conjecture. The idea of Ribet is to compare two Shimura curves such that two prime numbers $q$ and $q'$ dividing the discriminant for one and dividing the level for the other. Hence in weight 2 we could compare the Jacobians of corresponding Shimura curves. For most of our work, we extend the level-lowering part by replacing Ribet's method via Jacobians of certain Shimura curves with arguments using vanishing cycles on those curves (Proposition 5.7). This requires a study of Boutot-Carayol's version of Čerednik-Drinfel'd uniformization of Shimura curves [3]. With these results, the lifting of isomorphism $R_Q^D = \widehat{\mathbb{T}}_Q^D$ to $R_Q = \widehat{\mathbb{T}}_Q$ is carried out by applying the level-lowering (Proposition 5.9), and the numerical isomorphism criterion alluded to above.

Finally, using the local-to-global principle of Böckle [1], one can get rid of the set of auxiliary primes $Q$ and yields that the ramified minimal universal deformation $\rho_\emptyset$ is modular. That is, the canonical morphism $R_\emptyset \to \widehat{\mathbb{T}}_\emptyset$ is an actual isomorphism.

## 2. Deformation rings

Let $\mathbb{F}$ be a finite field of characteristic $p > 5$, $W = W(\mathbb{F})$ be the ring of Witt vectors with coefficients in $\mathbb{F}$, and $\mathcal{O}$ be a totally ramified extension of $W$ (hence its residue field is $\mathbb{F}$). Consider the continuous absolutely irreducible mod$p$ Galois representation $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$. We write $S$ to be the set of primes containing $p$, $\infty$ and the primes at which $\overline{\rho}$ is ramified, and $S' = S \setminus \{p\}$. Let $\mathrm{Ad}^0$ be the set of all trace zero two-by-two matrices over $\mathbb{F}$ with Galois action through $\overline{\rho}$ by conjugation.

Suppose that $\overline{\rho}$ is *modular* and satisfies the following conditions:

- The Serre weight $k := k(\overline{\rho})$ of $\overline{\rho}$ is greater than 2 and strictly less than $p$.
- $\det(\overline{\rho}) = \overline{\chi}_p$ is the mod$p$ cyclotomic character.
- $\mathrm{Ad}^0$ is absolutely irreducible.
- $\overline{\rho}$ is *semistable* at every primes in $S$.
- Moreover, $\overline{\rho}$ is *crystalline* and *ordinary* at $p$.

We refer to [14] for the existence of certain deformation rings parametrizing liftings of $\overline{\rho}$ with given local conditions.

Let $Q$ be a finite set of primes disjoint from $S$ such that for all $q \in Q$, $q \not\equiv \pm 1 \pmod{p}$ and $\overline{\rho}(\mathrm{Frob}_q)$ has eigenvalues with ratio $q$. Consider the following covariant deformation functor $\mathscr{D}_Q$ from the category of complete noetherian local $W$-algebras to the category of sets:

$$\begin{array}{rcl} \mathscr{D}_Q\colon\ \underline{\mathbf{CNL}_W} & \rightsquigarrow & \underline{\mathbf{Sets}} \\ (A,\varphi) & \rightsquigarrow & \{\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(A) \mid \rho \bmod \mathfrak{m}_A = \overline{\rho}\}/\sim, \end{array}$$

such that

(DC1) $\det \rho = \widetilde{\gamma}\chi_p^{k-1}$, where $\widetilde{\gamma}$ is the Teichmüller lifting of $S'$-ramified character $\gamma$ and $\chi_p$ is the $p$-adic cyclotomic character;

(DC2) $\rho$ is unramified outside $S \cup Q$;

(DC3) $\rho|_{I_\ell} = \langle \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \rangle$ for all $\ell \in S'$;

(DC4) for $\ell = p$, $\rho$ is ordinary and cristalline at $p$,

where $\rho_1 \sim \rho_2$ if and only if there exists $M \in \ker(\mathrm{GL}_2(A) \to \mathrm{GL}_2(\mathbb{F}))$ such that $\rho_1 = M^{-1}\rho_2 M$.

**Definition 2.1.** For any representation satisfying conditions (DC1)–(DC4), we will call it *minimally $S$-ramified.*

Since $\mathrm{Ad}^0$ is absolutely irreducible, then by the Schlessinger's criterion it is easy to see that the functor $\mathscr{D}_Q$ is pro-representable. We denote its universal couple by $(R_Q, \rho_Q)$.

*Remark* 2.2. There is *no* condition at any primes $q \in Q$.

More generally, for any subset $\alpha \subseteq Q$, we consider the closed subfunctor of $\mathscr{D}_Q$:

$$\mathscr{D}_Q^\alpha(A) = \{\rho\colon G_\mathbb{Q} \to \mathrm{GL}_2(A) \mid \rho \bmod \mathfrak{m}_A = \overline{\rho}\}/\sim$$

such that the conditions (DC1)–(DC4) hold and moreover

(DC5) $\rho|_{G_q} \sim \left(\begin{smallmatrix} \chi_p & * \\ 0 & 1 \end{smallmatrix}\right)$ for any $q \in \alpha$.

Since the functor $\mathscr{D}_Q^\alpha$ is relatively representable, hence the functor $\mathscr{D}_Q^\alpha$ is pro-representable. We denote the corresponding universal couple by $(R_Q^\alpha, \rho_Q^\alpha)$.

*Remark* 2.3. For $\alpha = Q$ and $D = \prod_{q \in Q} q$, we will write $R_Q^D$ instead of $R_Q^Q$. There is a sequence of natural surjections of local $W$-algebras $R_Q \twoheadrightarrow R_Q^\alpha \twoheadrightarrow R_Q^D$. If $Q = \emptyset$, we denote the corresponding universal couple by $(R_\emptyset, \rho_\emptyset)$, and call $R_\emptyset$ the *minimal deformation ring.*

## 2.1. The local conditions

Let $G_{S\cup Q}$ be the Galois group of the maximal extension of $\mathbb{Q}$ in $\overline{\mathbb{Q}}$ which is unramified outside $S \cup Q$. We introduce local conditions in order to define the Selmer group.

- For $v \in S'$, we let

$$\mathcal{L}_v = \mathrm{H}^1_{\mathrm{nr}}(G_v, \mathrm{Ad}^0) := \ker\left(\mathrm{H}^1(G_v, \mathrm{Ad}^0) \to \mathrm{H}^1(I_v, \mathrm{Ad}^0)\right).$$

- For $v = p$, we define

$$\mathcal{L}_p = \ker\left(\mathrm{H}^1(G_p, \mathrm{Ad}^0) \to \mathrm{H}^1(I_p, \mathrm{Ad}^0/Z)\right),$$

  where $Z$ consists of $\left(\begin{smallmatrix} 0 & * \\ 0 & 0 \end{smallmatrix}\right)$.

- For $v \in Q$, $\mathcal{L}_v$ is spanned by the 1-cocycles class given by

$$g(\sigma_v) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad g(\tau_v) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

modulo 1-coboundaries, where $\sigma_v$ and $\tau_v$ generate the same quotient of $G_v$ and satisfy $\sigma_v \tau_v \sigma_v^{-1} = \tau_v^{p_v}$.

Let $\mathcal{L}$ be the collection of these local conditions, and define the *Selmer group* to be

$$\mathrm{H}^1_{\mathcal{L}}(G_{S\cup Q}, \mathrm{Ad}^0) := \ker \left( \mathrm{H}^1(G_{S\cup Q}, \mathrm{Ad}^0) \to \bigoplus_{v \in S\cup Q} \mathrm{H}^1(G_v, \mathrm{Ad}^0)/\mathcal{L}_v \right).$$

The proof of the following proposition is routine.

**Proposition 2.4.** [15, §26] *Let $\mathfrak{m}_Q$ be the maximal ideal of $R_Q^D$. We have an isomorphism of $\mathbb{F}$-vector space:*

$$\mathrm{H}^1_{\mathcal{L}}(G_{S\cup Q}, \mathrm{Ad}^0) \simeq \mathrm{Hom}(\mathfrak{m}_Q/(p, \mathfrak{m}_Q^2), \mathbb{F}).$$

By [14, Proposition 21], there exists a finite set of primes $Q = \{q_2, \ldots, q_{2m}\}$ of odd cardinality such that for each $q \in Q$, $q \not\equiv \pm 1 \bmod p$, $\mathrm{Tr}\,\overline{\rho}(\mathrm{Frob}_q) = \pm(q+1)$, and such that

$$\mathrm{H}^1_{\mathcal{L}}(G_{S\cup Q}, \mathrm{Ad}^0) = 0.$$

Thus we have $\mathfrak{m}_Q = pR_Q^D$ by Proposition 2.4; therefore, $R_Q^D/pR_Q^D = \mathbb{F} = W/pW$ and by Nakayama's lemma the structure morphism $W \to R_Q^D$ is surjective. Since $k \le p-1$, the result of Diamond-Taylor [10] shows that there is a $p$-adic modular lifting of $\overline{\rho}$ which arises from a specialization of the universal representation $\rho_Q^Q \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R_Q^D)$ which gives a surjection from $R_Q^D$ to $W$. Hence the structure morphism is also injective, and we prove:

**Corollary 2.5.** *Let $2 \le k < p$. Then we have isomorphisms of local $W$-algebras*

$$W \simeq R_Q^D.$$

## 3. Some preliminaries

### 3.1. Theory of vanishing cycles

In this section we follow the presentation of Illusie [12]. Let $X$ be the proper semi-stable curve over $S = \mathrm{Spec}\,\mathbb{Z}_q$. For any constructible $\mathbb{Z}_p$-sheaves $\mathscr{F}$ on $X/S$, we have the *exact sequence of specialization* [12, §1.6]

$$\cdots \longrightarrow \mathrm{H}^0(X_{\overline{s}}, \mathrm{R}^1\,\Phi(\mathscr{F}))(1) \longrightarrow \mathrm{H}^2(X_{\overline{s}}, \mathrm{R}^0\,\Psi(\mathscr{F}))(1) \stackrel{\mathrm{sp}(1)}{\longrightarrow} \mathrm{H}^2(X_{\overline{\eta}}, \mathscr{F})(1) \longrightarrow 0,$$

where $\mathrm{R}^\bullet\,\Psi(\mathscr{F})$ (resp. $\mathrm{R}^\bullet\,\Phi(\mathscr{F})$) are the sheaves of *vanishing cycles* (resp. *nearby cycles*). We define $\mathbb{X}(\mathscr{F})$ to be:

$$\mathbb{X}(\mathscr{F}) := \ker\left(\bigoplus_{x\in\Sigma_1}(\mathrm{R}^1\,\Phi\mathscr{F})_x(1) \to \ker(\mathrm{sp})(1)\right),$$

where $\Sigma_1$ is the set of singular points of $X_{\overline{s}}$. Let $Y = X_{\overline{s}}$. The *cospecialization exact sequence* [12, §1.6] is

$$\cdots \longrightarrow \bigoplus_{x\in\Sigma_1}\mathrm{H}^1_x(Y,\mathrm{R}^1\,\Psi(\mathscr{F})) \xrightarrow{\beta'} \mathrm{H}^1(Y,\mathrm{R}^1\,\Psi(\mathscr{F})) \longrightarrow \cdots .$$

We let $\check{\mathbb{X}}(\mathscr{F}) := \mathrm{im}(\beta') \subset \mathrm{H}^1(X_{\overline{s}},\mathscr{F})$.

If $\mu$ is the normalization map of $\mu\colon \widetilde{Y} \to Y$ over $\overline{\mathbb{F}}_q$, we define the sheaf $\mathscr{G}$ on $Y$ by the exact sequence of sheaves:

$$0 \to \mathscr{F} \to \mu_*\mu^*\mathscr{F} \to \mathscr{G} \to 0.$$

Hence we obtain the following exact sequence

$$0 \longrightarrow \mathrm{H}^0(Y,\mathscr{F}) \longrightarrow \mathrm{H}^0(Y,\mu_*\mu^*\mathscr{F}) \xrightarrow{\theta} \mathrm{H}^0(Y,\mathscr{G}) \longrightarrow \cdots,$$

and we define

$$\check{\mathbb{Y}}_q(\mathscr{F}) := \mathrm{H}^0(Y,\mathscr{G})/\,\mathrm{im}(\theta),$$

and

$$\mathbb{Y}_q(\mathscr{F}) := \ker\left(\bigoplus_{x\in\Sigma_1}(\mathrm{R}^1\,\Phi\mathscr{F})_x(1) \to \ker(\mathrm{sp})(1)\right).$$

The monodromy pairing yields an injective map $\lambda_q\colon \mathbb{Y}_q(\mathscr{F}) \to \check{\mathbb{Y}}_q(\mathscr{F})$.

**Proposition 3.1.** $\check{\mathbb{X}}(\mathscr{F}) \simeq \mathrm{H}^0(Y,\mathscr{G})/\theta(\mathrm{H}^0(Y,\mu_*\mu^*\mathscr{F}))$.

*Proof.* Since the normalization map $\mu$ is finite, we can identify $\bigoplus_{x\in\Sigma_1}\mathrm{H}^1_x(Y,\mathrm{R}\,\Psi(\mathscr{F}))$ with $\mathrm{H}^0(Y,\mathscr{G})$. By Illusie [12, §1.5], we have the following "diagramme des 9" over $Y_{(x)}$ for the inclusions $i_x\colon \{x\} \hookrightarrow Y_{(x)}$ and $j_x\colon U_x = Y_{(x)} \setminus \{x\} \hookrightarrow Y_{(x)}$

$$\begin{array}{ccccc}
i_{x,*}\,\mathrm{R}\,i_x^!\mathscr{F} & \longrightarrow & \mathscr{F} & \longrightarrow & \mathrm{R}\,j_{x,*}j_x^*\mathscr{F} \\
\downarrow & & \downarrow & & \| \\
\mu_*\mu^*i_{x,*}\,\mathrm{R}\,i_x^!\mathscr{F} & \longrightarrow & \mu_*\mu^*\mathscr{F} & \longrightarrow & \mathrm{R}\,j_{x,*}j_x^*\mathscr{F} \\
\downarrow & & \downarrow & & \downarrow \\
\mathscr{G} & = & \mathscr{G} & \longrightarrow & 0
\end{array}$$

The composite

$$\gamma_x\colon \mathrm{H}^0(Y_{(x)}, \mu_*\mu^*\mathscr{F}) \to \mathrm{H}^0(Y_{(x)}, \mathrm{R}\, j_*j^*\mathscr{F}) \to \mathrm{H}^1(Y_{(x)}, i_{x,*}\, \mathrm{R}\, i_x^!\mathscr{F}) \simeq \mathrm{H}^1_x(Y, \mathscr{F})$$

is negative of the composite map

$$\gamma'_x\colon \mathrm{H}^0(Y_{(x)}, \mu_*\mu^*\mathscr{F}) \to \mathrm{H}^0(Y_{(x)}, \mathscr{G}) \to \mathrm{H}^1(Y_{(x)}, i_{x,*}\, \mathrm{R}\, i_x^!\mathscr{F}) \simeq \mathrm{H}^1_x(Y, \mathscr{F}),$$

by [12, Lemma 5.4]. In particular, the image of $\partial_x\colon \mathrm{H}^0(U_x, \mathscr{F}) \simeq \mathrm{H}^0(Y_{(x)}, \mathrm{R}\, j_{x,*}j_x^*\mathscr{F}) \to \mathrm{H}^1_x(Y, \mathscr{F})$ is the same as that of $\tau_x\colon \mathrm{H}^0(Y_{(x)}, \mathscr{G}) \to \mathrm{H}^1(Y_{(x)}, i_{x,*}\, \mathrm{R}\, i^!\mathscr{F}) \simeq \mathrm{H}^1_x(Y, \mathscr{F})$. The following commutative diagram

$$\begin{array}{ccc}
\mathrm{H}^0(U_x, \mathrm{R}\,\Psi(\mathscr{F})) & \xrightarrow{\ \partial'_x\ } & \mathrm{H}^1_x(Y, \mathrm{R}\,\Psi(\mathscr{F})) \\
\simeq \big\uparrow & & \big\downarrow \beta_x^{-1} \\
\mathrm{H}^0(U_x, \mathscr{F}) & \xrightarrow[\ \partial_x\ ]{} & \mathrm{H}^1_x(Y, \mathscr{F})
\end{array}$$

implies that the image of $\beta_x^{-1}$ is also the same as that of $\partial_x$. Thus, $\beta_x^{-1}$ and $\tau_x$ have the same image.

Consider the following commutative diagram

$$\begin{array}{ccccccc}
\mathrm{H}^0_{\Sigma_1}(Y, \mathscr{G}) & \longrightarrow & \bigoplus_{x\in\Sigma_1} \mathrm{H}^0_x(Y, \mathscr{G}) & \xrightarrow[(1)]{\simeq} & \bigoplus_{x\in\Sigma_1} \mathrm{H}^1_x(Y, \mathrm{R}\,\Psi(\mathscr{F})) & \longrightarrow & \mathrm{H}^1_{\Sigma_1}(Y, \mathrm{R}\,\Psi(\mathscr{F})) \\
\tau \big\downarrow & & \big\downarrow \bigoplus \tau_x & & \bigoplus \beta_x^{-1} \big\downarrow & & \big\downarrow \beta' \\
\mathrm{H}^1(Y, \mathscr{F}) & \xleftarrow[\ \sigma\ ]{} & \bigoplus_{x\in\Sigma_1} \mathrm{H}^1_x(Y, \mathscr{F}) & & \bigoplus_{x\in\Sigma_1} \mathrm{H}^1_x(Y, \mathscr{F}) & \xrightarrow[\ \sigma\ ]{} & \mathrm{H}^1(Y, \mathscr{F})
\end{array}$$

where (1) is given by the identification $\bigoplus_{x\in\Sigma_1} \mathrm{H}^1_x(Y, \mathrm{R}\,\Psi(\mathscr{F})) \simeq \mathrm{H}^0(Y, \mathscr{G})$. Thus we have $\mathrm{im}(\beta') = \mathrm{im}(\tau)$ and this is the same as $\mathrm{coker}(\theta)$. □

*Remark* 3.2. Note that $\mathbb{X}(\mathscr{F}) \subset \bigoplus_{x\in\Sigma_1}(\mathrm{R}^1\,\Phi\mathscr{F}_x)(1)$, so it is torsion-free. Thus we have the following exact sequence

$$0 \to \mathrm{H}^1(X_{\overline{s}}, \mathscr{F}) \otimes \mathbb{Z}/p^i\mathbb{Z} \to \mathrm{H}^1(X_{\overline{\eta}}, \mathscr{F}) \otimes \mathbb{Z}/p^i\mathbb{Z} \to \mathbb{X}(\mathscr{F})(-1) \otimes \mathbb{Z}/p^i\mathbb{Z} \to 0.$$

Similarly applying the snake lemma to the multiplication by $p^i$, we obtain

$$0 \to \check{\mathbb{X}}(\mathscr{F})[p^i] \to \Phi_q[p^i] \to \mathbb{X}(\mathscr{F}) \otimes \mathbb{Z}/p^i\mathbb{Z} \to \check{\mathbb{X}}(\mathscr{F}) \otimes \mathbb{Z}/p^i\mathbb{Z} \to \Phi_q \otimes \mathbb{Z}/p^i\mathbb{Z} \to 0.$$

Let $A$ be a $\mathbb{Z}_p$-algebra, and let $\Lambda_k(A) = \mathrm{Sym}^{k-2}\,\mathbb{Z}_p^2 \otimes_{\mathbb{Z}_p} A$. If we take $k-2 < p$ and $\mathscr{F} = \Lambda_k(\mathbb{Z}_p)$, we see that $\check{\mathbb{X}}(\Lambda_k(\mathbb{Z}_p))$ has no $p$-torsion since $\Lambda_k(\mathbb{Z}_p)$ is irreducible.

### 3.2. Shimura curves and Hecke correspondences

We review some properties of Shimura curves, following Buzzard [4]. We then take up the study of the reduction modulo a prime of the Shimura curves. As a preparation for our later work, we shall also present the Hecke correspondences.

Let $B$ be an indefinite quaternion algebra over $\mathbb{Q}$, and $S$ be the set of places where it ramifies. Let $D = \prod_{\ell \in S} \ell$, and let $M$ be a square-free integer prime to $D$. Let $\mathcal{O}_B$ be a maximal order of $B$. Consider an open compact subgroup $\Gamma$ of $\widehat{\mathcal{O}}_B^\times$ of level $M$ with determinant 1. Fix a prime $r$ not dividing $MD$. Let $\Gamma_0 = \Gamma \cap \widehat{\Gamma}_0^D(r)$ and $\Gamma_1 = \Gamma \cap \widehat{\Gamma}_1^D(r)$. We study here the reduction modulo $r$ of the Shimura curves $X^D(\Gamma_i)$ for $i = 0, 1$.

We fix an isomorphism $\phi_r \colon \mathcal{O}_B \otimes \mathbb{Z}_r \simeq \mathrm{M}_2(\mathbb{Z}_r)$, and let $e$ be the idempotent in $\mathcal{O}_B/r\mathcal{O}_B$ corresponding to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ by $\phi_r$. Following Buzzard [4], we define a $\Gamma_0(r)$-structure (resp. $\Gamma_1(r)$-structure) on a false elliptic curve $A$ as a finite flat group scheme $K_1$ of rank $r$ inside $(1-e)A[r]$ (resp. a Drinfel'd generator of this subgroup). For rigidification, we also introduce a $\Gamma$-level $\overline{\nu}$ structure on $A$, that is, a full level structure $\nu$ of level $N$ taken modulo $\Gamma$.

By [4, Corollary 4.2], the moduli problem on $\mathbb{Z}_r$-schemes $S \mapsto \{$isomorphism classes of $(A, \iota, \overline{\nu}, K_1)_{/S}\}$ is representable by a proper $\mathbb{Z}_r$-scheme which we denote by $X^D(\Gamma_0)$. This moduli problem is isomorphic to the problem $(A, \iota, \overline{\nu}, C)$ where $C$ is an isotropic subgroup of $A[r]$ of order $r^2$. (See the paragraph after Definition 3.1 of [4].) There is a universal triple $(A^u, \iota^u, K_1^u)$ defined over $X^D(\Gamma_0)$. Recall the following result in [4, Theorem 4.7]:

**Proposition 3.3.** (i) *The scheme $X^D(\Gamma_0)$ is proper over $\mathbb{Z}_r$.*

(ii) *It is semistable over $\mathbb{Z}_r$, i.e., regular, and smooth away from the supersingular points in characteristic $r$, with strictly henselian local ring at such a geometric point $\mathbb{Z}_r^{\mathrm{ur}}[\![X, Y]\!]/(XY - r)$; moreover, there are exactly two smooth irreducible components, $X^m$ and $X^e$, in the special fiber; they can be described as the Zariski closure of the locus $X^{m,0}$ where $K_1$ is of multiplicative type, resp. of $X^{e,0}$ where $K_1$ is étale.*

(iii) *The map $\pi \colon X^D(\Gamma_0) \to X^D(\Gamma)$ forgetting the $\Gamma_0$-structure is finite and flat.*

Following [11, Proposition 3.3.6] and using Tate-Oort theory, it is easy to prove:

**Proposition 3.4.** (i) *The model $X^D(\Gamma_1)$ of $X^D(\Gamma_1)_{\mathbb{Q}_r}$ is regular and flat over $\mathbb{Z}_r$.*

(ii) *The map $\pi_{10} \colon X^D(\Gamma_1) \to X^D(\Gamma_0)$ is finite flat; the special fiber of $X^D(\Gamma_1)$ is a divisor with normal crossings, with exactly two irreducible components $Y^e = \pi_{10}^{-1}(X^e)$ and $Y^m = \pi_{10}^{-1}(X^m)$ with multiplicity 1 and $r-1$ respectively, whose underlying reduced subschemes are smooth.*

(iii) *The two components cross (transversally) at the supersingular points and nowhere else.*

The curve $X^D(\Gamma_1)$ is a fine moduli space for triples $x = (A, \overline{\nu}, P)$, where, if $D > 1$, $A$ is a false elliptic curve with a level $\Gamma_1$ structure $\overline{\nu}$ and an $\mathcal{O}_B$-stable group scheme $K_1$ of rank $r$ inside $(1-e)A[r]$ and a Drinfel'd generator $P$ of $K_1$, and, if $D = 1$, $A$ is a generalized elliptic curve and a generator $P$ of a cyclic subgroup $K_1$ of order $r$. There is a universal triple $(A^u, \iota^u, K_1^u)$ defined over $X^D(\Gamma_1)$; let $f\colon A^u \to X^D(\Gamma_1)$. Similarly, $X^D(\Gamma_1 \cap \widehat{\Gamma}_0^D(\ell))$ classifies $(x, C)$ where $x$ is a triple as above and $C$ is an isotropic subgroup of order $\ell^2$ in $A[\ell]$.

For $\ell \nmid DMpr$, we have two degeneracy maps $\alpha_\ell$ and $\beta_\ell$ from $X^D(\Gamma_1 \cap \widehat{\Gamma}_0^D(\ell))$ to $X^D(\Gamma_1)$.

$$\begin{array}{ccc} & X^D(\Gamma_1 \cap \widehat{\Gamma}_0^D(\ell)) & \\ {}^{\alpha_\ell}\swarrow & & \searrow^{\beta_\ell} \\ X^D(\Gamma_1) & & X^D(\Gamma_1) \end{array}$$

They are defined by $\alpha_\ell((x, C)) = x$ and $\beta_\ell((x, C)) = (\phi_* x)$ where $\phi\colon A \to A/C$ denotes the quotient map and $\phi_* x = (A/C, \phi_*\overline{\nu}, \phi_* P)$.

If we are given a lisse sheaf $\mathscr{F}$ on $X^D(\Gamma_1)$ with a morphism $A_\ell\colon \beta_\ell^*\mathscr{F} \to \alpha_\ell^*\mathscr{F}$ over $X^D(\Gamma_1 \cap \widehat{\Gamma}_0^D(\ell))$, we can define Hecke correspondence acting on the pair $(X^D(\Gamma_1), \mathscr{F})$. By contravariant functoriality, it induces an endomorphism of $\mathrm{H}^\bullet(X^D(\Gamma_1), \mathscr{F})$ given by $T_\ell := \alpha_{\ell,*} \circ A_{\ell,*} \circ \beta_\ell^*$.

In our situation, we take $\mathscr{F} = \mathrm{Sym}^{k-2}\,\mathrm{R}\, f_*\mathbb{Z}_p$. The group $\Gamma_1$ acts on $\mathrm{Sym}^{k-2}\,\mathbb{Z}_p^2$ from the left by its $p$-component. We recall that the lisse sheaf associated to the corresponding representation of the fundamental group of $X^D(\Gamma_1)$ is $\mathrm{Sym}^{k-2}\,\mathrm{R}\, f_*\mathbb{Z}_p$. Let us consider the morphism $A_\ell\colon \beta_\ell^*\mathscr{F} \to \alpha_\ell^*\mathscr{F}$ induced by the left action of $\left(1, \ldots, 1, \left(\begin{smallmatrix} 1 & 0 \\ 0 & \ell \end{smallmatrix}\right), 1, \ldots, 1\right)$ on $\mathrm{Sym}^{k-2}\,\mathbb{Z}_p^2$. Note that this action being through the $p$-component is trivial if $\ell \neq p$. We define the $\ell^{\mathrm{th}}$ Hecke correspondence $T_\ell$ as $t_\ell(A_\ell)$.

For a false elliptic curve $A$, $(\mathbb{Z}/\ell\mathbb{Z})^\times$ acts on $A[\ell]$ by multiplication. We thus have an action of $(\mathbb{Z}/\ell\mathbb{Z})^\times$ on $\Gamma_1$-level structures on $A$. We let $\langle a\rangle(A, K_1, P) = (A, K_1, aP)$ as an endomorphism of $X^D(\Gamma_1)$.

These operators all commute with each other. We let $\mathbb{T}(\Gamma_1)$ denote the $\mathbb{Z}$-algebra generated by $T_\ell$ for all $\ell \nmid MDp$ and the diamond operators.

### 3.3. Nearby Cycles and monodromy

For our application to descent from an auxiliary level group $\widehat{\Gamma}_1^D(r)$ to a level prime to $r$, we shall need Langlands-Deligne-Carayol theorem on the compatibility between local and global Langlands correspondence for $B^\times$. For this purpose, we will consider a regular scheme $X = X^D(\Gamma_1)$ flat of finite type over $\mathbb{Z}_r$ with smooth generic fiber $X_\eta$ and special fiber $X_s$ in this section.

Let $\mathscr{F}$ be a $\mathbb{Z}/p^m$-module or a lisse sheaf of $\mathbb{Q}_p$-vector space over $X$. The special fiber is assumed to be étale locally one of the following types:

(1) *$X_s$ is smooth.* This corresponds to a neighbourhood of a point in $Y^m$ not in $Y^e$. Then the regularity of $X$ implies $X$ is smooth. In this case, $\mathrm{R}^q\,\Psi\mathscr{F} = 0$ for $q > 0$, and $\mathrm{R}^q\,\Psi\mathscr{F} = \mathscr{F}$.

(2) *$X$ is of the form* $\operatorname{Spec}\mathbb{Z}_r[A, B^{\pm 1}]/(A^{r-1}B - r)$. This corresponds to a neighbourhood of a point in $Y^e$ not in $Y^m$. Let $X' = \operatorname{Spec}\mathbb{Z}_r[a, b^{\pm 1}]/(a^{r-1} - r)$, and define a map $\pi\colon X' \to X$ via the embedding

$$\begin{array}{ccc} \mathbb{Z}_r[A, B^{\pm 1}]/(A^{r-1}B - r) & \to & \mathbb{Z}_r[a, b^{\pm 1}]/(a^{r-1} - r) \\ (A, B) & \mapsto & (ab^{-1}, b^{r-1}) \end{array}$$

The morphism $\pi$ is étale with Galois group $(\mathbb{Z}/r\mathbb{Z})^\times \simeq \mu_{r-1}$ where $\zeta \in \mu_{r-1}$ acts by multiplying both $a$ and $b$ by $\zeta$. The special fiber $X_s$ is a non-reduced divisor with multiplicity $r - 1$; the associated reduced divisor $X_{s,\mathrm{red}}$ is defined by $A = 0$ and is smooth; it is isomorphic to $\operatorname{Spec}\mathbb{F}_r[B^{\pm 1}]$ which we view as $(\mathbb{G}_m)_{\mathbb{F}_r}$.

We first compute the vanishing cycle $\mathrm{R}^q\,\Psi\mathscr{F}$ in the étale neighbourhood of $X'$ of $X$. Let $\mathcal{O} = \mathbb{Z}_r[\sqrt[r-1]{r}] = \mathbb{Z}_r[a]/(a^{r-1} - r)$. Then we write

$$X' = \operatorname{Spec}\mathcal{O} \times_{\operatorname{Spec}\mathbb{Z}_r} Y \tag{3.1}$$

where $Y = \operatorname{Spec}\mathbb{Z}_r[b^{\pm 1}]$. The second factor is smooth over $\mathbb{Z}_r$, hence $(\mathrm{R}^q\,\Psi\mathscr{F})_{X'}$ is the pullback from $\mathrm{R}^q\,\Psi\mathscr{F}$ for the finite flat morphism $\operatorname{Spec}\mathcal{O} \to \operatorname{Spec}\mathbb{Z}_r$. The morphism is of relative dimension zero, hence $\mathrm{R}^q\,\Psi\mathscr{F} = 0$ for $q > 0$. Similarly, $(\mathrm{R}^0\,\Psi\mathscr{F})_{\operatorname{Spec}\mathcal{O}}$ is the pull-back of $\operatorname{Spec}\mathcal{O} \to \operatorname{Spec}\mathbb{Z}_r$ and $(\mathrm{R}^0\,\Psi\mathscr{F})_{\operatorname{Spec}\mathbb{Z}_r} = \mathscr{F}$; hence $(\mathrm{R}^0\,\Psi\mathscr{F})_{\operatorname{Spec}\mathcal{O}} \simeq \mathscr{F}^{r-1}$ as $\mathbb{Z}/p^m$-modules or $\mathbb{Q}_p$-vector space. Since the inertia group of $\mathcal{O}$ over $\mathbb{Z}_r$ acting on $(\mathrm{R}^0\,\Psi\mathscr{F})_{\operatorname{Spec}\mathcal{O}}$ by $\mu_{r-1}$ transitively, we thus have $(\mathrm{R}^0\,\Psi\mathscr{F})_{\operatorname{Spec}\mathcal{O}}$ is the group algebra $\mathscr{F}[\mu_{r-1}]$.

It follows that $\mathrm{R}^q\,\Psi\mathscr{F} = 0$ for $q > 0$, and that $\mathrm{R}^0\,\Psi\mathscr{F}$ is a lisse $p$-adic sheaf of rank $r - 1$ on $X_{s,\mathrm{red}}$ that becomes constant over $X'_{s,\mathrm{red}}$. Moreover, since $\operatorname{Gal}(X'/X)$ acts as inertia group on the first factor of (3.1), one sees that the canonical action of $\operatorname{Gal}(X'/X)$ on $(\mathrm{R}^0\,\Psi\mathscr{F})_{X'_{s,\mathrm{red}}}$ identifies the latter with the group algebra $\mathscr{F}[\operatorname{Gal}(X'/X)]$. It follows that

$$\mathrm{R}^0\,\Psi\mathscr{F} \xrightarrow{\simeq} \pi_*\mathscr{F}.$$

The inertia group $\mu_{r-1}$ acts on $\mathrm{R}^0\,\Psi\mathscr{F}$, and we have seen that the lift of this action to $X'_{s,\mathrm{red}}$ coincides with the action of $\operatorname{Gal}(X'/X)$. We write

$$\mathrm{R}^0\,\Psi\mathscr{F} = \bigoplus_{\chi} \mathrm{R}^0\,\Psi\mathscr{F}[\chi],$$

the decomposition with respect to characters of the inertia group. We write $L = \mathrm{R}^0\,\Psi\mathscr{F}$, and $L[\chi]$ for the rank one local system $\mathrm{R}^0\,\Psi\mathscr{F}[\chi]$. Let $\chi_0$ denote the trivial character. Consider the embedding

$$i\colon X_{s,\mathrm{red}} = (\mathbb{G}_m)_{\mathbb{F}_r} \hookrightarrow \mathbb{A}^1 = \operatorname{Spec}\mathbb{F}_r[B]$$

as the complement of the origin $B = 0$. The morphism $\pi$ is totally ramified along $B = 0$. It follows that

$$\text{(3.2)}\qquad \begin{aligned} &\mathrm{R}^0\,i_*L[\chi] = i_!L[\chi]; \quad \mathrm{R}^q\,i_*L[\chi] = 0, \quad q > 0 \quad (\chi \neq \chi_0);\\ &\mathrm{R}^0\,i_*L[\chi_0] = \mathscr{F}; \qquad \mathrm{R}^q\,i_*L[\chi_0] = 0, \quad q > 0. \end{aligned}$$

More generally, suppose $X = \operatorname{Spec} R[A, B^{\pm 1}]/(A^{r-1}B - r)$, where $R$ is a smooth $\mathbb{Z}_r$-algebra of finite type. Then $X$ is the fiber product

$$X = \operatorname{Spec} R \times_{\operatorname{Spec}\mathbb{Z}_r} \operatorname{Spec}\mathbb{Z}_r[A, B^{\pm 1}]/(A^{r-1}B - r),$$

where the first factor is smooth. We define

$$i_X = 1 \times i\colon X_{s,\mathrm{red}} = \operatorname{Spec} R \times_{\operatorname{Spec}\mathbb{Z}_r} \operatorname{Spec}(\mathbb{G}_m)_{\mathbb{F}_r} \to \operatorname{Spec} R \times_{\operatorname{Spec}\mathbb{Z}_r} \operatorname{Spec}\mathbb{F}_r[B].$$

Let $X_2$ denote the second factor above and let $\mathrm{pr}_2$ denote the projection $X \to X_2$. We see that

$$\mathrm{R}^q\,\Psi\mathscr{F} = \mathrm{pr}_2^*\,\mathrm{R}^q\,\Psi_{X_2}\mathscr{F},$$

where $\mathrm{R}\,\Psi_{X_2}$ denotes the vanishing cycle sheaves for the map from $X_2$ to $\operatorname{Spec}\mathbb{Z}_r$. In particular, $\mathrm{R}^q\,\Psi\mathscr{F} = 0$ for $q > 0$, while $\mathrm{R}^q\,\Psi\mathscr{F}$ breaks up under the action of the inertia subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}_r/\mathbb{Q}_r)$ as the sum of rank one local system $L[\chi]$: $L[\chi_0] = \mathscr{F}$, whereas $L[\chi]$ for nontrivial $\chi$ satisfies the analogue of (3.2):

$$\mathrm{R}^0\,i_{X,*}L[\chi] = i_{X,!}L[\chi]; \quad \mathrm{R}^q\,i_{X,*}L[\chi] = 0,\ q > 0 \quad (\chi \neq \chi_0).$$

(3) *$X$ is of the form* $\operatorname{Spec} R[A, B]/(A^{r-1}B - r)$*, where $R$ is a smooth $\mathbb{Z}_r$-algebra of finite type.* This corresponds to a neighbourhood of a point in $Y^e \cap Y^m$. We will calculate the stalks of $\mathrm{R}^q\,\Psi\mathscr{F}$ at a geometric point $\overline{x}$ of the singular locus $X_{\mathrm{sing}}$ of the special fiber defined by $A = B = 0$. In this case, we simply have $\mathrm{R}^q\,\Psi\mathscr{F} = 0$, $q > 1$; $(\mathrm{R}^0\,\Psi\mathscr{F})_{\overline{x}} = \mathscr{F}$, $(\mathrm{R}^1\,\Psi\mathscr{F})_{\overline{x}} = \mathscr{F}(-1)$, with trivial action of the inertia group on $\mathscr{F}$, $(-1)$ denoting Tate twist.

We write $Y^a = Y^m \cap Y^e$. Let $i_m\colon Y^m \to X^D(\Gamma_1)$, $i_e\colon Y^e \to X^D(\Gamma_1)$, and $i_a\colon Y^a \to X^D(\Gamma_1)$ be the natural maps. Let $Y_0^e$ denote the complement of $Y^a$ in $Y^e$, and let $j_e\colon (Y_0^e)_{\mathrm{red}} \to (Y^e)_{\mathrm{red}}$ be the open immersion. Then the vanishing cycle sheaves $\mathrm{R}^q\,\Psi\mathscr{F}$ are calculated as follows:

**Proposition 3.5.** *Let $I$ denote the inertia subgroup of* $\mathrm{Gal}(\overline{\mathbb{Q}}_r/\mathbb{Q}_r)$. *Then the action of $I$ on $\mathrm{R}\,\Psi^q\mathscr{F}$ factors through the map to $(\mathbb{Z}/r\mathbb{Z})^\times$ (which we identify with $\mu_{r-1}(\mathbb{Z}_r)$ by Teichmüller lifting) given by the action on $\mathbb{Q}[\zeta_r]$. For a character $\chi$ of $\mu_{r-1}$, let $[\chi]$ denote the $\chi$-isotypic component, and let $\chi_0$ denote the trivial character. Then*

(i) $\mathrm{R}^0\,\Psi\mathscr{F}[\chi_0] = \mathscr{F}$.

(ii) $\mathrm{R}^1\,\Psi\mathscr{F} = \mathscr{F}[\chi_0]$ *is a rank one local system supported on $Y^a$, locally isomorphic at any point of $Y^a$ to $\mathscr{F}(-1)$.*

(iii) *For $\chi \neq \chi_0$, $\mathrm{R}^0\,\Psi\mathscr{F}$ is the extension by zero of a rank one lisse sheaf $L[\chi]$ supported on $Y_0^m$. Moreover, the natural map $i_{m,!}\,\mathrm{R}^0\,\Psi\mathscr{F}[\chi] \to \mathrm{R}\,i_{m,*}\Psi^0\mathscr{F}[\chi]$ is a quasi-isomorphism.*

(iv) $\mathrm{R}^q\,\Psi\mathscr{F} = 0$ *for $q > 1$.*

*Proof.* Everything follows from the cases (1)–(3) discussed above except the global triviality of $\mathrm{R}^0\,\Psi\mathscr{F}[\chi_0]$. But there is always an injection $\mathscr{F} \to \mathrm{R}^0\,\Psi\mathscr{F}[\chi_0]$, so (i) follows from the fact that all stalks of $\mathrm{R}^0\,\Psi\mathscr{F}[\chi_0]$ are one-dimensional. □

Since the same vanishing cycle sheaves are concentrated in two degrees, the vanishing cycle spectral sequence degenerates into a long exact sequence

$$\begin{aligned}\cdots \to \mathrm{H}^i(X^D(\Gamma_1)_{\overline{s}}, \mathrm{R}^0\,\Psi\mathscr{F}) &\to \mathrm{H}^i(X^D(\Gamma_1)_{\overline{\eta}}, \mathscr{F})\\ \to \mathrm{H}^{i-1}(X^D(\Gamma_1)_{\overline{s}}, \mathrm{R}^1\,\Psi\mathscr{F}) &\to \mathrm{H}^{i+1}(X^D(\Gamma_1)_{\overline{s}}, \mathrm{R}^0\,\Psi\mathscr{F}) \to \cdots .\end{aligned}$$

Using Proposition 3.5(ii), we rewrite this

$$\begin{aligned}\cdots \to \mathrm{H}^i(X^D(\Gamma_1)_{\overline{s}}, \mathrm{R}^0\,\Psi\mathscr{F}) &\to \mathrm{H}^i(X^D(\Gamma_1)_{\overline{\eta}}, \mathscr{F})\\ &\to \mathrm{H}^{i-1}((Y^a)_{\mathrm{red}}, \mathrm{R}^1\,\Psi\mathscr{F}[\chi_0]) \to \cdots .\end{aligned} \tag{3.3}$$

We deduce from (i) and (iii) of Proposition 3.5 that the first term in turn is calculated by a long exact sequence

$$\begin{aligned}\cdots \to \mathrm{H}^{i-1}((Y^a)_{\mathrm{red}}, \mathscr{F}) &\to \mathrm{H}^i(X^D(\Gamma_1)_{\overline{\eta}}, \mathrm{R}^0\,\Psi\mathscr{F})\\ \to \mathrm{H}^i(Y^m, \mathscr{F}) \oplus \mathrm{H}^i((Y^e)_{\mathrm{red}}, \mathscr{F}) &\oplus \bigoplus_{\chi\neq\chi_0} \mathrm{H}^i_c(Y_0^e, L[\chi]) \to \cdots .\end{aligned} \tag{3.4}$$

Here and in (3.3), we have replaced $Y^e$ and $Y^a$ by the associated reduced schemes, since the étale cohomology is insensitive to nilpotents.

The diamond operators act $X^D(\Gamma_1)_{\overline{\eta}}$ as well as on $Y^e$ and $Y^m$, and thus induce compatible actions on the spaces in the exact sequence (3.3) and (3.4). These are determined as follows:

**Lemma 3.6.** *The diamond operators* $\langle a\rangle$ *act on the outer terms of the exact sequence* (3.4) *as follows: The action acts via* $\chi$ *on* $L[\chi]$*, and acts trivially on* $\mathrm{H}^\bullet((Y^e)_{\mathrm{red}}, \mathscr{F})$*, on* $\mathrm{H}^\bullet(Y^a, \mathscr{F})$*, and on* $\mathrm{H}^{i-1}((Y^a)_{\mathrm{red}}, \mathrm{R}^1\,\Psi\mathscr{F}[\chi_0])$*.*

*Proof.* The diamond operators act trivially on $(Y^e)_{\mathrm{red}}$ and $(Y^a)_{\mathrm{red}}$, so it suffices to determine their action on $\mathrm{R}^0\,\Psi\mathscr{F}$ and $\mathrm{R}^1\,\Psi\mathscr{F}$.

For $\mathrm{R}^0\,\Psi\mathscr{F}$, by the discussion in (2), we see it suffices to determine the action of the diamond operators on $\mathrm{H}^0(\mathrm{Spec}(\overline{\mathbb{Q}}_r \otimes_{\mathbb{Q}_r} \mathbb{Q}_r[\zeta_r]), \mathscr{F})$, via the identification of $\mathbb{Q}_r[\zeta_r]$ with the generic fiber of $\mu_{r-1}$ and the latter with $C_R$ in (1)–(2). But the diamond operators on $\mu_{r-1}$ are tautologically given by the cyclotomic character.

For the action on $\mathrm{R}^1\,\Psi\mathscr{F}$, this is again local. But locally the discussion in (3) shows that $\mathrm{R}^1\,\Psi\mathscr{F}$ is a constant sheaf, so the triviality of the action of the diamond operators is clear. □

**Proposition 3.7.** *Suppose* $\chi \neq \chi_0$*, and denote by* ${}^{\langle\,\rangle=\chi}$ *the* $\chi$*-isotypic component for the action of the diamond operators. Then for any* $i$*, there is a canonical isomorphism of* $\mathrm{Gal}(\overline{\eta}/\eta)$*-modules*

$$\mathrm{H}^i(Y^m, \mathscr{F})^{\langle\,\rangle=\chi} \oplus \mathrm{H}^i_c((Y^e_0)_{\mathrm{red}}, L[\chi]) \xrightarrow{\simeq} \mathrm{H}^i(X^D(\Gamma_1)_{\overline{\eta}}, \mathscr{F})^{\langle\,\rangle=\chi}.$$

*Proof.* Indeed, in (3.4), the diamond operators act trivially on the term $\mathrm{H}^i((Y^a)_{\mathrm{red}}, \mathscr{F})$ and coincide with inertia on $L[\chi]$, inducing an isomorphism

$$\mathrm{H}^i(X^D(\Gamma_1)_{\overline{s}}, \mathrm{R}^0\,\Psi\mathscr{F})^{\langle\,\rangle=\chi} \xrightarrow{\simeq} \mathrm{H}^i(Y^m, \mathscr{F})^{\langle\,\rangle=\chi} \oplus \mathrm{H}^i_c(Y^e_0, L[\chi]).$$

Similarly, the diamond operators act trivially on the $\mathrm{H}^{i-1}((Y^a)_{\mathrm{red}}, \mathrm{R}^1\,\Psi\mathscr{F}[\chi_0])$ term in (3.4). □

## 4. Hecke rings and modular Galois representations

Let $N \geq 1$ be a square-free integer. We let $\Gamma_0(N)$ be the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ consisting of elements $\left(\begin{smallmatrix} a & b \\ Nc & d\end{smallmatrix}\right)$. For all $\ell \nmid N$, the Hecke operators $T_\ell$ acting on $S_k(\Gamma_0(N))$ generate a $\mathbb{Z}$-algebra $\mathbb{T}$. Let $f \in S_k(\Gamma_0(N))$ be an eigen cusp newform of weight $k \geq 2$. We fix a prime $p > 5$ not dividing $N$. Let $K$ be the $p$-adic field generated by the Fourier coefficients of $f$, let $\mathcal{O}$ be the ring of integers of $K$, and let $\mathbb{F}$ be its residue field. We consider the attached character $\lambda_f\colon \mathbb{T} \to K$ and assume that $f$ is *ordinary*; that is, $\lambda_f(T_p) \in \mathcal{O}_K^\times$. We are interested in the Galois representation $\rho_f\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(K)$ attached to $f$. Let $\mathfrak{p} = \ker(\lambda_f)$ and let $\mathfrak{m}_f$ be the unique maximal ideal of $\mathbb{T}$ containing $\mathfrak{p} + p\mathbb{T}$.

Suppose that

(H1) $2 \leq k < p$;

(H2) $\mathfrak{m}_f \subset \mathbb{T}$ is non-Eisenstein;

(H3) the residual representation $\overline{\rho}_f \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ is minimally ramified at primes dividing $pN$;

(H4) $\overline{\rho}_f|_{I_p} = \left(\begin{smallmatrix} \overline{\chi}_p^{k-1} & * \\ 0 & 1 \end{smallmatrix}\right)$ with $* \neq 0$.

Consider a finite set of primes $Q = \{q_2, \ldots, q_{2m}\}$ of odd cardinality such that $\overline{\rho}_f$ is unramified at primes in $Q$, $q_i \not\equiv \pm 1 \bmod p$ for $q_i \in Q$, and such that $\mathrm{Tr}(\overline{\rho}_f(\mathrm{Frob}_{q_i})) = \pm(q_i+1)$ for $q_i \in Q$. Let $D = \prod_{q \in Q} q$. We write $\widetilde{N} = pND$, and we fix a prime $q_1$ dividing $pN$.

For $0 \leq s \leq m$, let $Q_s = \{q_1, q_2, \ldots, q_{2m-2s}\}$ and let $B_s$ be the indefinite quaternion algebra ramified on $Q_s$. Let $D_s = \prod_{i=1}^{m-s} q_i$ and $M_s = \widetilde{N}/D_s$ and for $0 \leq s \leq m$. Choose an Eichler order $R_{M_s,D_s}$ of level $M_s$ in $B_s$. Denote the corresponding Shimura curve by $X^{D_s}(M_s)$.

Let $A$ be a $\mathbb{Z}_p$-algebra, and let $\Lambda_k(A) = \mathrm{Sym}^{k-2}\,\mathbb{Z}_p^2 \otimes_{\mathbb{Z}_p} A$. Notice that if $k < p$ and $A$ is a $\mathbb{Z}_p$-flat algebra, $\mathrm{H}^1(X^{D_s}(M_s), \Lambda_k(A))$ is a torsion-free $A$-module. If $A = \mathbb{Z}_p$, we simply write it $\Lambda_k$ or $\Lambda$. The $\ell^{\mathrm{th}}$-Hecke correspondence $T_\ell$ defines an endomorphism, still denoted by $T_\ell$, of $\mathrm{H}^1(X^{D_s}(M_s), \Lambda_k(A))$ for all $\ell \nmid \widetilde{N}$. The *Hecke algebra* $\mathbb{T}_Q^{D_s}$ is the $A$-algebra generated by these endomorphisms $T_\ell$ for all primes $\ell \nmid \widetilde{N}$. For $A = W$, we drop the $A$ in the notation and we simply write $\mathbb{T}_Q^{D_s}$. We also have the *minimal Hecke algebra* $\mathbb{T}_\emptyset$ generated over $W$ by Hecke operators $T_\ell$ on the corresponding modular curve $X(\widehat{\Gamma}_0(Np))$ for all primes $\ell$ such that $(\ell, \widetilde{N}) = 1$.

For any $0 \leq s \leq m$, we have obvious surjective $W$-algebra homomorphisms

$$\mathbb{T}_Q^{D_s} \to \mathbb{T}_\emptyset \to \mathbb{T}.$$

We let $\mathfrak{m}_Q$ be the preimage of $\mathfrak{m}_f$ under the map $\mathbb{T}_Q^{D_s} \to \mathbb{T}$, and denote the completion of the Hecke algebra $\mathbb{T}_Q^{D_s}$ at $\mathfrak{m}_Q$ by $\widehat{\mathbb{T}}_Q^{D_s}$. Note that for any $0 \leq s \leq m$, the Hecke algebra $\widehat{\mathbb{T}}_Q^{D_s}$ is finite flat over $W$.

**Lemma 4.1.** (i) *We have Galois representations*

$$\rho_{Q,\mathrm{mod}}^{D_s} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\widehat{\mathbb{T}}_Q^{D_s}) \quad \textit{(resp. } \rho_{\emptyset,\mathrm{mod}} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\widehat{\mathbb{T}}_{\mathfrak{m}_f})\textit{)}$$

*which are unramified outside $S \cup Q$ such that for $\ell \notin S \cup Q$,*

$$\mathrm{Tr}\,\rho_{Q,\mathrm{mod}}^{D_s}(\mathrm{Frob}_\ell) = T_\ell \quad \textit{and} \quad \mathrm{Tr}\,\rho_{\emptyset,\mathrm{mod}}(\mathrm{Frob}_\ell) = T_\ell.$$

*They arise by uniquely determined specializations of the universal representations*

$$\rho_Q^{D_s} \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R_Q^{D_s}) \quad \textit{(resp. } \rho_\emptyset \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R_\emptyset)\textit{).}$$

(ii) *They satisfy*

$$\rho_{Q,\mathrm{mod}}^{D_s} \in \mathscr{D}_{Q,\mathcal{O}}^{D_s}(\widehat{\mathbb{T}}_Q^{D_s}) \quad \textit{(resp. } \rho_{\emptyset,\mathrm{mod}} \in \mathscr{D}_{\mathcal{O}}(\widehat{\mathbb{T}}_{\mathfrak{m}_f})\textit{)}.$$

(iii) *The local $\mathcal{O}$-algebra homomorphisms defined by the universal property*

$$R_Q^{D_s} \otimes_W \mathcal{O} \to \widehat{\mathbb{T}}_Q^{D_s} \quad \textit{(resp. } R_\emptyset \otimes_W \mathcal{O} \to \widehat{\mathbb{T}}_{\mathfrak{m}_f}\textit{)}$$

*are surjective.*

*Proof.* By the irreducibility of the residual representations, we can apply the theorem of Carayol [8] and Nyssen [16] to construct a representation using pseudo-representations on $\widehat{\mathbb{T}}_Q^{D_s}$ and $\widehat{\mathbb{T}}_Q^{D_s} \otimes \mathbb{Q}$. By definition of the representation, it satisfies the local conditions; also by [8], the same holds for its integral structure.

According to Carayol [7], the representations $\rho_{Q,\mathrm{mod}}^{D_s} \mid G_q$ and $\rho_{\emptyset,\mathrm{mod}} \mid G_q$ are of the form

$$\pm \begin{pmatrix} \chi_p & * \\ 0 & 1 \end{pmatrix}.$$

Hence, the existence of the specialization map follows from the universal property of deformation ring $R_Q^{D_s}$. □

Lemma 4.1 says that we have a surjective $W$-homomorphism $R_Q^D \twoheadrightarrow \widehat{\mathbb{T}}_Q^D$, and since the algebra $\widehat{\mathbb{T}}_Q^D$ is finite flat over $W$, we see that the map $R_Q^D \twoheadrightarrow \widehat{\mathbb{T}}_Q^D$ is injective. Hence, by Corollary 2.5, we deduce the following:

**Proposition 4.2.** *Let $2 \le k < p$. Then for $D = \prod_{q \in Q} q$ we have isomorphisms of local $W$-algebras*

$$W \simeq R_Q^D \xrightarrow{\sim} \widehat{\mathbb{T}}_Q^D.$$

*We call such set $Q$ a* Khare-Ramakrishna system *or a* Khare-Ramakrishna set.

A new proof the following theorem, that will be deduced from Proposition 4.2, will be given in the next section.

**Theorem 4.3.** *Let $N$ be a square-free positive integer, $p > 5$ be a prime not dividing $N$, and $2 \le k < p$ be an integer. Suppose that $f \in S_k(\Gamma_0(N))$ is a cusp newform and let $\overline{\rho} = \overline{\rho}_f \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be the* mod$p$ *Galois representation attached to $f$. Assume $\overline{\rho}$ satisfies conditions* (H1)–(H4). *Then the universal deformation ring $R$ associated to $\overline{\rho}$ is canonically isomorphic to $\widehat{\mathbb{T}}_\emptyset$ ($\simeq \widehat{\mathbb{T}}$).*

*Remark* 4.4. The analogue result has been proven by Taylor-Wiles in weight 2 case using the Taylor-Wiles systems method; it has been generalized by Ramakrishna [17] following a similar method. In weight 2, it has been reproved by Khare; for higher weights, we will prove the theorem stated above by following the ideas of Khare.

## 4.1. The auxiliary prime $r$

For technical reasons — mostly to assure that the $q$-adic uniformizations occurring in our proof involve only torsion-free hyperbolic groups, we shall need to introduce an auxiliary prime $r$. Its goal is to get rid of the possible torsion of the various arithmetic groups that we consider without creating any new ramification in the modular Galois representations considered.

Let us consider the following condition on a prime $r$:

$$\begin{aligned}\mathrm{UR}(r) : r \nmid \widetilde{N}, r \not\equiv 1 \bmod p, &\text{ and the ratio of the eigenvalues of } \rho(\mathrm{Frob}_r) \\ &\text{is not congruent to 1 or } r^{\pm 1} \bmod p.\end{aligned}$$

The existence of a prime $r$ satisfying $\mathrm{UR}(r)$ follows from Čebotarev's density theorem.

We fix such a prime $r$ in the sequel. For any $0 \leq s \leq m$, let $\Gamma_1(M_s; r) = \widehat{R}^{\times}_{M_s,D_s} \cap \widehat{\Gamma}^{D_s}_1(r)$ and $\Gamma_0(M_s; r) = \widehat{R}^{\times}_{M_s,D_s} \cap \widehat{\Gamma}^{D_s}_0(r)$. Denote the corresponding Shimura curves by $X^{D_s}(M_s; r)$. Let $A$ be a $\mathbb{Z}_p$-algebra. Applying the Jacquet-Langlands correspondence, we denote by $\mathbb{T}^{rQ_s}_Q$ the *Hecke algebra* generated by Hecke correspondences $T_\ell$ on $\mathrm{H}^1(X^{D_s}(M_s; r), \Lambda_k(A))$ for all $\ell \nmid \widetilde{N}$ for primes $\ell \nmid \widetilde{N}$. Note that if $s = m$, we also have the *minimal Hecke algebra* $\mathbb{T}_\emptyset$ generated by Hecke operators $T_\ell$ on the corresponding moduli curve $X(\widehat{\Gamma}_0(Np) \cap \widehat{\Gamma}_1(r))$ for primes $\ell$ such that $(\ell, \widetilde{N}) = 1$.

To simplify our notations, we set $\Gamma_1 = \widehat{R}^{\times}_{M_s,D_s} \cap \widehat{\Gamma}^{D_s}_1(r)$ and $\Gamma_0 = \widehat{R}^{\times}_{M_s,D_s} \cap \widehat{\Gamma}^{D_s}_0(r)$, and let $X^{D_s}(\Gamma_1)$ and $X^{D_s}(\Gamma_0)$ be the corresponding Shimura curves for each $0 \leq s \leq m$. Let $\mathfrak{m}_{(r)} = \mathfrak{m} + (T_r - \alpha_r - \beta_r, rS_r - \alpha_r\beta_r)$. If $\pi'$ occurs in $\mathrm{H}^1(X^{D_s}(\Gamma_1), \Lambda)^{T_r\equiv\alpha_r}_{\mathfrak{m}_{(r)}}$ is special, then it occurs in $\mathrm{H}^1(X^{D_s}(\Gamma_0), \Lambda)^{T_r\equiv\alpha_r}_{\mathfrak{m}_{(r)}}$. By the weight monodromy conjecture for curves, the eigenvalues of $\mathrm{Frob}_r$ on $\rho_{\pi'} \subset \mathrm{H}^1(X^{D_s}(\Gamma_0), \Lambda)^{T_r\equiv\alpha_r}_{\mathfrak{m}_{(r)}}$ are of the form $\alpha'_r$, $r\alpha'_r$. (See also Carayol [7].) However, we have

$$\rho_\pi(\mathrm{Frob}_r) \sim \begin{pmatrix} \alpha_r & 0 \\ 0 & \beta_r \end{pmatrix}$$

which implies that $\alpha_r/\beta_r \equiv r^{\pm 1}$ modulo $p$, and deduces a contradiction. Hence, $\pi'_r$ belongs to the ramified principal series and there exists a non-trivial character $\chi$ of $(\mathbb{Z}/r\mathbb{Z})^\times$ such that the diamond operator acts on $\pi'_r$ by $\chi$. In particular, $\pi'$ occurs in $\mathrm{H}^1(X^{D_s}(\Gamma_1)_{\overline{\eta}}, \Lambda)^{\langle\rangle=\chi}_{\mathfrak{m}_{(r)}}$. Then by Proposition 3.7, we see that

$$\rho_{\pi'}|_{I_r} \sim \begin{pmatrix} 1 & 0 \\ 0 & \chi \end{pmatrix}.$$

Since $r \not\equiv 1 \bmod p$, this implies that $\overline{\rho}_{\pi'} = \overline{\rho}$ is also ramified at $r$ which is a contradiction. Therefore, $\rho_{\pi'}$ is unramified.

*Remark* 4.5. Since the associated Galois representations of the Hecke algebras $\widehat{\mathbb{T}}_Q^{rD_s}$ are unramified at $r$, we still have the surjective specialization maps of local $W$-algebras $R_Q^{D_s} \to \widehat{\mathbb{T}}_Q^{rD_s}$ for $0 \leq s \leq m$. Hence we will ignore the auxiliary prime $r$ in the sequel to reduce our notations.

## 5. Proof of Theorem 4.3

The strategy of the proof of Theorem 4.3 is to deduce from it from Proposition 4.2 using the *Numerical Criterion* and the *higher weight version of level-lowering method* [19] to compute change of $\eta$-invariants when we relax the conditions of $Q$ on the deformation and Hecke rings two primes at a time.

### 5.1. A numerical inequality

As in Wiles' method, the proof of modularity is based on a numerical inequality relating the length of a congruence module to the cardinality of a Selmer group. We recall a numerical criterion due to Lenstra which refines a result of Wiles.

**Theorem 5.1** (Numerical Criterion)**.** *Let $R, T \in \underline{\mathbf{CNL}}_W$. Suppose that $T$ is finite flat as $W$-module and $\phi\colon R \to T$ is a surjective local $W$-algebra homomorphism. Let $\pi\colon T \to W$ be a homomorphism of local $W$-algebras, and set $\Phi(R) = \ker(\pi\phi)/\ker(\pi\phi)^2$ and $\eta_T = \pi(\mathrm{Ann}_T(\ker(\pi)))$. Then we have the following:*

(i) $|W/\eta_T| \leq |\Phi(R)|$.

(ii) *Assume that $\eta_T$ is not zero. Then the following are equivalent:*

- *The equality $|W/\eta_T| = |\Phi(R)|$ is satisfied.*
- *The rings $R$ and $T$ are complete intersections and $\phi$ is an isomorphism.*

In our case, we let $\phi\colon R_Q^\alpha \to \widehat{\mathbb{T}}_Q^\alpha$ and $\pi\colon \widehat{\mathbb{T}}_Q^\alpha \to \widehat{\mathbb{T}}_Q^D$. For any prime $q \in Q$, let $t_q$ generate the unique $\mathbb{Z}_p$ quotient of $I_q$. Then $\rho_Q^{D_s}(t_q)$ is of the form

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

for some $x_q \in W \setminus \{0\}$, and $(x_q)$ does not depend on the choice of an integral model for $\rho_Q^D$. Note that $x_q \neq 0$. Indeed, if $x_q = 0$ then $\rho_Q^D$ is unramified at $q$, and as $R_Q^D \simeq \widehat{\mathbb{T}}_Q^D$ we have $\mathrm{Tr}\,\rho_Q^D(\mathrm{Frob}_q) = \pm(q+1)$ which contradicts the Ramanujan bound.

We obtain an upper bound for $\Phi(R_Q^\alpha)$ by identifying $\Phi(R_Q^\alpha)$ with the dual of Selmer group, and thus we have the following result which is proved in [14, Lemma 16]:

**Proposition 5.2.** *For any subset $\alpha \subset Q$, we have*

$$\left|\Phi(R_Q^\alpha)\right| \leq \prod_{q \in Q \setminus \alpha} |W/(x_q)|.$$

*Remark* 5.3. For any $\alpha \subset Q$, the invariant $\Phi(R_Q^\alpha)$ can be thought of as the cotangent space of the scheme $\mathrm{Spec}(R_Q^\alpha)$ at the point $\ker(\pi\phi)$.

## 5.2. Ribet's short exact sequence

### 5.2.1. Residual characteristic divides the level

Suppose that the prime $q$ does not divide the discriminant $D$ of the indefinite quaternion algebra $B_{/\mathbb{Q}}$. Let $\Gamma = \Gamma_q \Gamma^q = \widehat{\Gamma}_0(qM) \cap \widehat{\Gamma}_1^D(r)$ be an open compact subgroup of $\widehat{B}^\times$. We have defined the Shimura curves $X^D(qM; r)$ associated to $\Gamma$. Let $V_0(q) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}_q) \;\middle|\; c \equiv 0 \bmod q \right\}$.

Let $\mu$ be the normalization map for the special fiber $Y$ of $X = X^D(qM; r)$ over $\mathbb{Z}_q$. Let $\mathbb{T}_{qM;r}$ denote the Hecke algebra generated over $\mathbb{Z}_q$ by the endomorphisms $T_\ell$ $(\ell \nmid Mqr)$ of $\mathrm{H}^1(X^D(qM; r), \Lambda)$. We will write $\mathbb{X}_q(qM; r)$ (resp. $\check{\mathbb{X}}_q(qM; r)$) instead of $\mathbb{X}(\Lambda)$ (resp. $\check{\mathbb{X}}(\Lambda)$) in order to emphasis the level structure.

**Proposition 5.4.** *If $\mathfrak{m}$ is a non-Eisenstein maximal ideal of $\mathbb{T}_{qM;r}$, then we have:*

(i) $\mathrm{im}(\theta)_{\mathfrak{m}} = 0$, *and we have a canonical isomorphism* $\check{\mathbb{X}}_q(qM; r)_{\mathfrak{m}} \simeq \left(\bigoplus_{x \in \Sigma_1} \mathscr{G}_x\right)_{\mathfrak{m}}$.

(ii) *In the exact sequence of specialization, after localization at $\mathfrak{m}$, the map*

$$\bigoplus_{x \in \Sigma_1} \left(\mathrm{R}^1\,\Phi(\Lambda)\right)_x \to \ker(\mathrm{sp})$$

*is the zero map, i.e.,* $\mathbb{X}_q(qM; r)_{\mathfrak{m}} \simeq \left(\bigoplus_{x \in \Sigma_1} \left(\mathrm{R}^1\,\Phi(\Lambda)\right)_x\right)_{\mathfrak{m}}$.

(iii) *From* (i) *and* (ii), *we deduce the following*

$$\check{\mathbb{X}}_q(qM; r)_{\mathfrak{m}} \simeq \left(\bigoplus_{x \in \Sigma_1} \mathscr{G}_x\right)_{\mathfrak{m}} \overset{(1)}{\simeq} \left(\bigoplus_{x \in \Sigma_1} \left(\mathrm{R}^1\,\Phi(\Lambda)\right)_x\right)_{\mathfrak{m}} \simeq \mathbb{X}_q(qM; r)_{\mathfrak{m}},$$

*where* (1) *is induced from the monodromy logarithm $N_x$ at each $x$.*

*Proof.* (i) Note that the normalization map

$$\mu\colon X^D(M; r) \otimes \overline{\mathbb{F}}_q \sqcup X^D(M; r) \otimes \overline{\mathbb{F}}_q \to X^D(qM; r) \otimes \overline{\mathbb{F}}_q$$

is a finite morphism. We thus have an isomorphism

$$\mathrm{H}^0(X^D(qM; r) \otimes \overline{\mathbb{F}}_q, \mu_* \mu^* \mathscr{F}) \simeq \mathrm{H}^0(X^D(M; r) \otimes \overline{\mathbb{F}}_q, \mathscr{F})^2.$$

In Carayol [6, §2], we have that

$$\begin{aligned}\pi_0(X^D(M;r)\otimes\overline{\mathbb{F}}_q) &\simeq \mathbb{Q}_+^\times \setminus (\mathbb{A}^\infty)^\times / \operatorname{Nm}(\widehat{\Gamma}_0(M)\cap\widehat{\Gamma}_1^D(r)) \\ &= (\mathbb{Z}_{(q)})_+^\times \setminus (\mathbb{A}^{q\infty})^\times / \operatorname{Nm}(\widehat{\Gamma}_0(M)^q\cap\widehat{\Gamma}_1^D(r)^q),\end{aligned}$$

where we write $\mathbb{Z}_{(q)}$ for $\mathbb{Q}\cap\mathbb{Z}_q$. Then $G(\mathbb{A}^{q\infty})$ acts on the set of components through via reduced norm (see [6, §1.3]). But the maximal ideals in $\mathbb{T}_{qq'M}$ lying in the support of $\check{\mathbb{X}}$ must correspond to the one-dimensional automorphic representations, as cuspidal representations on quaternion groups admit infinite-dimensional components at almost every place, and thus do not factor through the norm (see [7, §4.4]). However, the automorphic representations in $\operatorname{im}(\theta)$ factor through the norm. By Proposition 3.1, we deduce $\check{\mathbb{X}}_q(qM;r)_{\mathfrak{m}} \simeq \big(\bigoplus_{x\in\Sigma_1}\mathscr{G}_x\big)_{\mathfrak{m}}$.

(ii) Since $\mathscr{G}$ concentrates at points, its cohomology groups vanish in degree greater than one. Hence, we also have an isomorphism induced from the normalization map $\mu$:

$$\mathrm{H}^2(X^D(qM;r)\otimes\overline{\mathbb{F}}_q,\mathscr{F}) \simeq \mathrm{H}^2(X^D(qM;r)\otimes\overline{\mathbb{F}}_q,\mu_*\mu^*\mathscr{F}).$$

On the other hand, we may regard the latter group as $\mathrm{H}^2(X^D(M;r)\otimes\overline{\mathbb{F}}_q,\mathscr{F})^2$, and this is Poincaré dual to the group $\mathrm{H}^0(X^D(M;r)\otimes\overline{\mathbb{F}}_q,\check{\mathscr{F}}(1))^2$. Using similar analysis in (i), the second point follows as before.

(iii) The homomorphisms $N_x$ are isomorphisms for any regular model for $X$ over $\mathbb{Z}_q$. The third point (iii) follows from this together with the first and the second assertions. □

### 5.2.2. Ribet's short exact sequence

We let $X = X^D(M;r)$ and assume that $qq' \mid D$. Let $\Gamma\subset\widehat{B}^\times$ be the group of level $M$ defining $X$. We write $\Gamma = \Gamma_q\Gamma^q$, where $\Gamma_q = \mathcal{O}_{B_q}^\times$ and $\Gamma^q$ is an open compact subgroup of $B^\times(\mathbb{A}^q)$. We also insist that $\Gamma_{q'} = \mathcal{O}_{B_{q'}}^\times$. Let $\mathrm{GL}_2(\mathbb{Q}_q)_+$ (resp. $\mathrm{GL}_2(\mathbb{Q}_q)_+$) be the subset of elements in $\mathrm{GL}_2(\mathbb{Q}_q)$ whose reduced norm has even (resp. odd) valuation. Let $B'$ be the definite quaternion of discriminant $D/q$ obtaining from $B$ by exchanging the local invariants at $q$ and $\infty$. The Čerednik-Drinfel'd uniformization theorem gives a description of the dual graph $\mathcal{G}$ of the special fiber $Y$ of $X$ at $q$.

The set of edges of $\mathcal{G}$ is $\operatorname{Ed}(\mathcal{G}) = V_0(q)\Gamma^q\setminus\widehat{B}'^\times/B'^\times$. Let us introduce

$$\mathscr{V}_+ := \mathrm{GL}_2(\mathbb{Q}_q)_+\setminus(\mathrm{PGL}_2(\mathbb{Q}_q)_+/\mathrm{PGL}_2(\mathbb{Z}_q)\times Z_\Gamma)$$

and

$$\mathscr{V}_- := \mathrm{GL}_2(\mathbb{Q}_q)_+\setminus(\mathrm{PGL}_2(\mathbb{Q}_q)_-/\mathrm{PGL}_2(\mathbb{Z}_q)\times Z_\Gamma)$$

where $Z_\Gamma = \Gamma^q\setminus\widehat{B}'^\times/B'^\times$. Then the set of vertices of $\mathcal{G}$ is $\operatorname{Ver}(\mathcal{G}) = \mathscr{V}_+\sqcup\mathscr{V}_-$.

Define $\mathscr{V} := \mathrm{GL}_2(\mathcal{O}_q)\Gamma^q\setminus\widehat{B}'^\times/B'^\times = \mathrm{GL}_2(\mathcal{O}_q)\setminus Z_\Gamma$. Note that we have two degeneracy maps $\alpha$, $\beta$ from $\operatorname{Ed}(\mathcal{G})$ to $\mathscr{V}$ corresponding to the inclusion of $V_0(q)\Gamma^q$ into $\mathrm{GL}_2(\mathcal{O}_q)\Gamma^q$

and the conjugation by $W_q = \left(\begin{smallmatrix} 1 & 0 \\ 0 & q \end{smallmatrix}\right)$. We have bijections between $\mathscr{V}$ and $\mathscr{V}_+$, and $\mathscr{V}$ and $\mathscr{V}_-$. Each edge $e$ connects $\alpha(e)$ in $\mathscr{F}_+$ to $\beta(e)$ in $\mathscr{V}_-$. In fact, we have

$$\begin{aligned} 1\colon \mathscr{V} &\to \mathscr{V}_+, && [x] \mapsto (1, x) \\ W_q\colon \mathscr{V} &\to \mathscr{V}_-, && [x] \mapsto (W_q, W_q x). \end{aligned}$$

To simplify the notations and state Carayol's result in the sequel, we make the following assumptions: Let $B_1$ be an indefinite quaternion algebra over $\mathbb{Q}$ with discriminant $D_1$, $p \nmid D_1$, $M_1 \geq 4$, and let $X^{D_1}(M_1; r) = G(\mathbb{Q})\backslash G(\mathbb{A}^\infty) \times \mathcal{H}_\infty / Z(\mathbb{A}^\infty)\Xi$ where $\Xi = \widehat{\Gamma}_0^{D_1}(M_1) \cap \widehat{\Gamma}_1^{D_1}(r) \subset G(\mathbb{A}^\infty)$. We denote by $K$ the restricted product of $(B_1 \otimes \mathbb{Q}_v)^\times$ for $v \neq p$.

**Proposition 5.5.** [5,6] *Let $\Sigma_{M_1;r}$ be the set of supersingular points of $X^{D_1}(M_1; r)$ modulo $p$. Then the group $K \times \mathbb{Q}_p^\times$ acts transitively on $\Sigma_{M_1;r}$. For each $x \in \Sigma_{M_1;r}$, the stabilizer of $x$ is conjugate in $K \times \mathbb{Q}_p^\times$ to $\overline{Z(\mathbb{Q})}G'(\mathbb{Q})$ where $G' = B_2^\times$ obtaining from $B_1$ by changing the local invariants at $p$ and $\infty$ and $\overline{Z(\mathbb{Q})}$ is the closure of $Z(\mathbb{Q})$ in $Z(\mathbb{A}^\infty)$.*

Apply these information to our cases with $M_1 = qq'M$, $D_1 = D/qq'$, $p = q'$, $B = B_1$ and $B' = B_2$. Let $\Xi = \widehat{\Gamma}_0(qq'M) \cap \widehat{\Gamma}_1(r)$; consider the modulo $q'$ reduction of $X^{D/qq'}(qq'M; r)$, notice that $\left(\widehat{\Gamma}_0(qq'M) \cap \widehat{\Gamma}_1(r)\right)^{q'} \times \mathcal{O}_{B'_{q'}}^\times \simeq V_0(q)\Gamma^q$, and this gives a one-to-one correspondence between $\mathrm{Ed}(\mathcal{G})$ and the set of singular points of $X^{D/qq'}(qq'M; r)$ modulo $q'$.

$$\begin{aligned} &\mathrm{Ed}(\mathcal{G}) : \text{the edges of the dual graph of } X^D(M; r) \bmod q \\ \Longleftrightarrow\ &\Sigma_{qq'M;r} : \text{singular points of } X^{D/qq'}(qq'M; r) \bmod q'. \end{aligned}$$

Similarly, for vertices of the dual graph of the Shimura curve $X^D(M; r)$ we use Carayol's formula for $\mathscr{L} = \widehat{\Gamma}_0(q'M) \cap \widehat{\Gamma}_1(r)$. We find that

$$\left(\widehat{\Gamma}_0(q'M) \cap \widehat{\Gamma}_1(r)\right)^{q'} \times \mathcal{O}_{B'_{q'}}^\times \simeq \mathrm{GL}_2(\mathcal{O}_q)\Gamma^q.$$

The correspondence provides a bijection for $\mathscr{V}_?$ $(? = \emptyset, +, -)$:

$$\mathscr{V}_?\ (? = \emptyset, +, -) \iff \Sigma_{q'M;r}\text{: singular points of } X^{D/qq'}(q'M; r) \bmod q'.$$

Therefore, the map $1_*$ (resp. $W_{q,*}$) will correspond to $\alpha$ (resp. $\beta$).

The number of irreducible components of the normalization is equal to the number of the vertices of the dual graph of the special fiber. Hence, for the lisse sheaf $\Lambda$ we let $\zeta$ be the composition of two Hecke-equivariant maps (1) and (2)

$$\begin{aligned} \zeta : \mathrm{H}^0(X_{\overline{s}}, \mu_* \mu^* \Lambda) &\overset{(1)}{\simeq} 1_* \left( \bigoplus_{y \in \Sigma_{q'M;r}} \mathscr{G}_y \right) \oplus W_{q,*} \left( \bigoplus_{y \in \Sigma_{q'M;r}} \mathscr{G}_y \right) \\ &\overset{(2)}{\twoheadrightarrow} 1_* \check{\mathbb{X}}_{q'}(q'M; r) \oplus W_{q,*} \check{\mathbb{X}}_{q'}(q'M; r), \end{aligned}$$

where (1) follows from Proposition 5.4 and (2) follows Proposition 3.1. We let $J = \ker(\zeta)$. Following from Proposition 5.4, $\mathrm{H}^0(X_{\overline{s}}, \mu_*\mu^*\Lambda)/J \simeq \mathrm{im}(\zeta)$ is a $\mathbb{T}_{qq'M;r}$-module. The Hecke-equivariant injection $\mathrm{H}^0(Y,\Lambda) \to \mathrm{H}^0(Y, \mu_*\mu^*\Lambda)$ induces an injection of $\mathrm{H}^0(Y,\Lambda)/(J \cap \mathrm{H}^0(Y,\Lambda)$ into $\mathrm{H}^0(X_{\overline{s}}, \mu_*\mu^*\Lambda)/J$. Hence, this map induces a $\mathbb{T}_{qq'M;r}$-module structure on $\mathrm{H}^0(Y,\Lambda)/(J \cap \mathrm{H}^0(Y,\Lambda))$.

**Lemma 5.6.** *We have* $\left(\mathrm{H}^0(Y,\Lambda)/J \cap \mathrm{H}^0(Y,\Lambda)\right)_{\mathfrak{m}} = 0$.

*Proof.* Since the restriction of $\Lambda$ to each irreducible component of $Y$ is constant, $\mathrm{H}^0(Y,\Lambda)$ is isomorphic to a direct sum of $\Lambda_y$'s, each corresponding to a connected component of $Y$. Hence, as the connected components of $X \otimes \overline{\mathbb{Q}}$ are defined over $\mathbb{Q}^{\mathrm{ab}}$, the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action on $\mathrm{H}^0(X \otimes \overline{\mathbb{Q}}, \Lambda)$ factors through $\mathrm{Gal}(\overline{\mathbb{Q}}^{\mathrm{ab}}/\mathbb{Q})$. This gives rise to a reducible representation. So $\mathrm{H}^0(Y,\Lambda)/J \cap \mathrm{H}^0(Y,\Lambda)$ is Eisenstein. □

By Proposition 3.1, we also have a surjective homomorphism of $\mathbb{Z}_p$-modules

$$\eta\colon\ \mathrm{H}^0(Y, \mathscr{G}) \to \check{\mathbb{X}}_q(qq'M;r).$$

Let $\mathscr{J} = \ker(\eta)$. On $\mathrm{H}^0(Y,\mathscr{G})/\mathscr{J}$, we have a $\widehat{\mathbb{T}}_{qq'M;r}$-module structure. By the previous lemma, we may identify $\left(\mathrm{H}^0(Y, \mu_*\mu^*\Lambda)/J\right)_{\mathfrak{m}}$ with $\left(\mathrm{H}^0(Y, \mu_*\mu^*\Lambda)/\theta^{-1}\mathscr{J}\right)_{\mathfrak{m}}$. By the definition of $\check{\mathbb{Y}}_q(q'M;r)$, we have the following exact sequence of $\widehat{\mathbb{T}}_{qq'M;r}$-modules

$$0 \to \left(\mathrm{H}^0(Y, \mu_*\mu^*\Lambda)/J\right)_{\mathfrak{m}} \to \left(\mathrm{H}^0(Y,\mathscr{G})/\mathscr{J}\right)_{\mathfrak{m}} \to \check{\mathbb{Y}}_q(q'M;r) \to 0.$$

Finally, we obtain the following exact sequence

$$0 \longrightarrow \check{\mathbb{X}}_{q'}(q'M;r)^2_{\mathfrak{m}} \overset{1_*\oplus W_{q,*}}{\longrightarrow} \check{\mathbb{X}}_{q'}(qq'M;r)_{\mathfrak{m}} \longrightarrow \check{\mathbb{Y}}_q(q'M;r) \longrightarrow 0.$$

We see $\mathrm{H}^2(Y,\Lambda) \simeq \mathrm{H}^2(Y, \mu_*\mu^*\Lambda)$ and $\mathrm{H}^2(Y, \mu_*\mu^*\Lambda) = \mathrm{H}^2(\widetilde{Y}, \mu^*\Lambda)$. Since the components of $Y$ correspond to a disjoint union of two copies of $\Sigma_{q'M;r}$, $\mathrm{H}^2(\widetilde{Y}, \mu^*\Lambda)$ is isomorphic as a $\mathbb{Z}_p$-module to the direct sum of two copies of $\bigoplus_{x\in\Sigma_{q'M;r}} \mathrm{R}\,\Phi(\Lambda)_x$. Therefore, we have two injective homomorphisms of $\mathbb{Z}_p$-modules

$$f_1\colon \mathbb{X}_{q'}(q'M;r)^2 \hookrightarrow \left(\bigoplus_{x\in\Sigma_{q'M;r}} \mathrm{R}\,\Phi(\Lambda)_x\right)^2(1) \simeq \mathrm{H}^0(Y,\Lambda)(1)$$

and

$$f_2\colon \mathbb{X}_{q'}(qq'M;r) \hookrightarrow \bigoplus_{x\in\Sigma_{qq'M;r}} \mathrm{R}\,\Phi(\Lambda)_x(1) \simeq \bigoplus_{x\in\Sigma_1} \mathrm{R}\,\Phi(\Lambda)_x(1).$$

Using similar argument above, we finally get the following exact sequence

$$0 \to \mathbb{Y}_q(q'M;r)_{\mathfrak{m}} \to \mathbb{X}_{q'}(qq'M;r)_{\mathfrak{m}} \to \mathbb{X}_{q'}(q'M;r)^2_{\mathfrak{m}} \to 0.$$

Since monodromy pairings are Hecke-equivariant, we deduce:

**Proposition 5.7.** *We have the following commutative diagram of $\widehat{\mathbb{T}}_{qq'M;r}$-modules where the rows are exact and the vertical maps come from monodromy pairings:*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \check{\mathbb{X}}_{q'}(q'M;r)^2_{\mathfrak{m}} & \xrightarrow{1_*\oplus W_{q,*}} & \check{\mathbb{X}}_{q'}(qq'M;r)_{\mathfrak{m}} & \longrightarrow & \check{\mathbb{Y}}_q(M;r)_{\mathfrak{m}} & \longrightarrow & 0 \\ & & \uparrow & & \uparrow{\scriptstyle \lambda'_{q'}} & & \uparrow{\scriptstyle \lambda_q} & & \\ 0 & \longleftarrow & \mathbb{X}_{q'}(q'M;r)^2_{\mathfrak{m}} & \xleftarrow[1^*\oplus W_q^*]{} & \mathbb{X}_{q'}(qq'M;r)_{\mathfrak{m}} & \xleftarrow[\iota]{} & \mathbb{Y}_q(M;r)_{\mathfrak{m}} & \longleftarrow & 0 \end{array}$$

In our application, we will ignore the auxiliary prime $r$ in the proposition, since it is immaterial in our calculations below (cf. Remark 4.5).

For a $\mathbb{T}$-module $L$, we will denote by $L_{\mathfrak{m},W}$ the module $L$ localized at $\mathfrak{m}$ and then tensored with $W$.

## 5.3. The level-lowering

We apply Ribet's short exact sequence to our descending induction by increasing $s$ via posing:

$$\begin{array}{ll} q' = q_{2m-2s} & q = q_{2m-2s-1} \\ \mathscr{P} = \ker\left(\pi\colon \widehat{\mathbb{T}}_Q^{D_s} \to \widehat{\mathbb{T}}_Q^{D} \simeq W\right) & \\ \xi\colon \mathbb{Y}_q(M)_{\mathfrak{m}}[\mathscr{P}] \to \check{\mathbb{Y}}_q(M)_{\mathfrak{m}} & \xi^* \text{ dual of } \xi \\ \xi'\colon \mathbb{X}_{q'}(qq'M)_{\mathfrak{m}}[\mathscr{P}] \to \check{\mathbb{X}}_{q'}(qq'M)_{\mathfrak{m}} & \xi'^* \text{ dual of } \xi'. \end{array}$$

The maps $\xi^*\xi$ and $\xi'^*\xi'$ commute with the $W[G_{\mathbb{Q}}]$-action and can be regarded as given by multiplication by elements of $W$. We denote the corresponding ideals of $W$ by $(\xi^*\xi)$ and $(\xi'^*\xi')$.

Let $\mathscr{L} := \mathbb{X}_{q'}(qq'M)_{\mathfrak{m}}[\mathscr{P}]$ and $\mathscr{L}' := \mathbb{Y}_q(M)_{\mathfrak{m}}[\mathscr{P}]$. Via the map $\iota$ in the Ribet's short exact sequence, we may identify $\mathscr{L}$ with $\mathscr{L}'$: Since $\mathscr{L} \cap \mathbb{Y}_q(M)_{\mathfrak{m},W}$ is of rank one and the Ribet's short exact sequence shows that $\mathscr{L}$ in $\mathbb{X}_{q'}(qM)^2_{\mathfrak{m},W}$ is finite. In fact, it is zero since $\mathbb{X}_{q'}(qM)^2_{\mathfrak{m},W}$ is torsion-free by the assumption $k < p$.

The following lemma and proposition are crucial in the proofs of the Theorem 4.3.

**Lemma 5.8.** (i) $|\operatorname{coker}(\lambda_q)| = |W/(x_q)|$ *for any* $q \in Q$.

(ii) *For any* $\ell \mid Np$, $|\operatorname{coker}(\lambda_\ell)| = 1$.

(iii) $[\mathbb{X}_{q'}(qq'M)_{\mathfrak{m}} : \mathscr{L}] = |x_{q'}|^{-1}$.

(iv) $[\mathbb{Y}_q(M)_{\mathfrak{m}} : \mathscr{L}] = 1$.

*Proof.* (i) We see that the monodromy pairing acts from $\mathbb{Y}_q(M)_{\mathfrak{m}}$ to $\check{\mathbb{Y}}_q(M)_{\mathfrak{m}}$ as $\sigma - 1$ for $\sigma \in I$. Hence $|\operatorname{coker}(\lambda_q)| = |W/(x_q)|$ for any $q \in Q$.

(ii) Since we assume the minimality in our deformation of representations, for any $\ell \mid N$ the generator of the pro-$p$ part of the inertia at $\ell$ acts by $\left(\begin{smallmatrix} 1 & x_\ell \\ 0 & 1 \end{smallmatrix}\right)$ with $x_\ell \in W^\times$ so that $\mathrm{coker}(\lambda_\ell)$ is trivial. For $\ell = p$, since we assume $\overline{\rho}|_{I_p}$ is not split, the generator of the pro-$p$ part of the inertia at $p$ acts by $\begin{pmatrix} \chi_p^{k-1} & x_p \\ 0 & 1 \end{pmatrix}$ with $x_p \in W^\times$. Hence $\mathrm{coker}(\lambda_p)$ is trivial as well.

(iii) We have a surjection $\xi'_*\colon \mathbb{X}_{q'}(qq'M)_\mathfrak{m} \to \mathbb{X}_{q'}(qq'M)_\mathfrak{m}[\mathscr{P}]$. Using the monodromy parings, we have

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbb{X}_{q'}(qq'M)_\mathfrak{m} & \longrightarrow & \mathrm{Hom}_W(\mathbb{X}_{q'}(qq'M)_\mathfrak{m}, W) & \longrightarrow & \mathrm{coker}(\lambda'_{q'}) & \longrightarrow & 0 \\ & & \downarrow \xi'_* & & \downarrow f & & \downarrow g & & \\ 0 & \longrightarrow & \mathbb{X}_{q'}(qq'M)_\mathfrak{m}[\mathscr{P}] & \longrightarrow & \mathrm{Hom}_W(\mathbb{X}_{q'}(qq'M)_\mathfrak{m}[\mathscr{P}], W) & \longrightarrow & \mathrm{coker}(\lambda_{q'}) & \longrightarrow & 0 \end{array}$$

This implies the order of $\mathrm{coker}(g)$ equals to the order of $\mathrm{coker}(f)$. But the order of $\mathrm{coker}(f)$ is the order of the torsion subgroup of $\mathrm{coker}((\mathbb{X}_{q'}(qq'M)_\mathfrak{m}[\mathscr{P}] \to (\mathbb{X}_{q'}(qq'M)_\mathfrak{m})$. This derives that $\left|\mathrm{coker}(\mathrm{coker}(\lambda'_{q'}) \to \mathrm{coker}(\lambda_{q'}))\right| = [\mathbb{X}_{q'}(qq'M)_\mathfrak{m} : \mathscr{L}]$. Recall that $\mathbb{X}_{q'}(qq'M)_\mathfrak{m}$ is self-dual; thus $\mathrm{coker}(\lambda'_{q'}) = 1$ and the result is deduced from (ii).

(iv) The fourth assertion is proved analogously. □

**Proposition 5.9.** *We have*

$$(\xi'^*\xi') = (\xi^*\xi)(x_q x_{q'}),$$

*where when $s = m - 1$ we declare $x_q$ to be a unit.*

*Proof.* Take a generator $\alpha$ of $\mathbb{X}_{q'}(qq'M)_\mathfrak{m}[\mathscr{P}]$ and let $(\tau) = \langle \alpha, \alpha \rangle$. Let $\beta = \iota^{-1}(\alpha)$. By the adjoint property of $\xi'$ with respect to the monodromy pairing, we have

$$(\xi'^*\xi')(\langle \alpha, \alpha \rangle) = (\xi'^*\xi')(x_q) = (\langle \xi'(\alpha), \xi'(\alpha) \rangle) = \left([\mathbb{X}_{q'}(qq'M)_\mathfrak{m} : \mathscr{L}]^{1/d}\right)^2 \cdot (\tau)$$

where $d = \mathrm{rank}_{\mathbb{Z}_p}(W)$. Similarly, we have $(\langle \beta, \beta \rangle) = (\tau)$ and

$$(\xi^*\xi)(x_q) = (\langle \xi'(\alpha), \xi'(\alpha) \rangle) = \left([\mathbb{Y}_q(M)_\mathfrak{m} : \mathscr{L}]^{1/d}\right)^2 \cdot (\tau).$$

The result now follows easily from the previous lemma. □

## 5.4. End of proof

Using Proposition 5.9 inductively, we obtain

$$(\xi_m^*\xi_m) \subset \prod_{q \in Q} x_q.$$

Let $\mathscr{L} := \mathbb{Y}[\ker(\pi)]$, where $\mathbb{Y} = \mathbb{Y}_{q_1}(N)$. By Lemma 5.8, we see that

$$[\mathbb{Y} : \mathscr{L}] = 1,$$

and this deduces a map $\xi_m$ from $\mathscr{L}$ to $\mathbb{Y}$ with torsion-free cokernel. Let $x$ be a generator of the free rank one $W$-module $\mathscr{L}$. By the adjoint property of monodromy pairing and Lemma 5.8, we see that $\xi_r^*(x)$ generates $\mathscr{L}$ and that

$$(\langle \xi_m(x), \xi_m(x)\rangle) = ((\xi_m^*\xi_m)\langle x, x\rangle) = (\xi_m^*\xi_m).$$

Define $I := \mathrm{Ann}_{\mathbb{T}_Q}(\ker(\pi))$ and $\eta_{\mathbb{T}_Q} := \pi(I)$. Since $\mathbb{Y}/\mathbb{Y}[I] \simeq \mathrm{Hom}_W(\mathscr{L}, W)$ we have

$$\mathbb{Y}/(\mathbb{Y}[\ker(\pi)] + \mathbb{Y}[I]) \simeq \mathrm{coker}(\mathscr{L} \to \mathrm{Hom}_W(\mathscr{L}, W))$$

as $W/\eta_{\mathbb{T}_Q}$-modules. Note that $\mathbb{Y}/(\mathbb{Y}[\ker(\pi)] + \mathbb{Y}[I])$ is also annihilated by $\eta_{\mathbb{T}_Q}$. Since we have

$$|W/\eta_{\mathbb{T}_Q}| \geq |\mathbb{Y}/(\mathbb{Y}[\ker(\pi)] + \mathbb{Y}[I])| = |\mathrm{coker}(\mathscr{L} \to \mathrm{Hom}_W(\mathscr{L}, W))|,$$

this implies $\eta_{\mathbb{T}_Q} \subset (\xi_m^*\xi_m) \subset (\prod_{q\in Q} x_q)$.

Applying the Numerical Criterion (Theorem 5.1) and Proposition 5.2, we thus deduce the isomorphism

$$R_Q \simeq \widehat{\mathbb{T}}_Q.$$

Finally, an argument of Böckle [2, Theorem 1] implies that:

$$R_\emptyset \simeq \widehat{\mathbb{T}}_\emptyset$$

is an isomorphism of complete intersection rings. This completes the proof.

## Acknowledgments

The author is deeply indebted to Jacques Tilouine, who suggested the problem, for many enlightening discussions. Moreover, he is very thankful to R. Livné for indicating the thesis of L. Yang [22], and to A. Mokrane who pointed out the article of L. Illusie [12]. He also thanks J. Yu of NTU, and C.-F. Yu of Academia Sinica for their hospitality.

## References

[1] G. Böckle, *A local-to-global principle for deformations of Galois representations*, J. Reine Angew. Math. **509** (1999), 199–236.

[2] ———, *On the isomorphism $R_\phi \to T_\phi$*, Invent. Math. **154** (2003), 218–222.

[3] J.-F. Boutot and H. Carayol, *Uniformisation p-adique des courbes de Shimura: les théorèmes de Čerednik et de Drinfel'd*, Astérisque **7** (1992), 45–158.

[4] K. Buzzard, *Integral models of certain Shimura curves*, Duke Math. J. **87** (1997), no. 3, 591–612.

[5] H. Carayol, *Sur la mauvaise réduction des courbes de Shimura*, C. R. Acad. Sci. Paris Sér. I Math. **296** (1983), no. 13, 557–560.

[6] ———, *Sur la mauvaise réduction des courbes de Shimura*, Compositio Math. **59** (1986), no. 2, 151–230.

[7] ———, *Sur les représentations $l$-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. École Norm. Sup. (4) **19** (1986), no. 3, 409–468.

[8] ———, *Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet*, in *$p$-adic monodromy and the Birch and Swinnerton-Dyer conjecture* (Boston, MA, 1991), 213–237, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994.

[9] F. Diamond, *On deformation rings and Hecke rings*, Ann. of Math. (2) **144** (1996), no. 1, 137–166.

[10] F. Diamond and R. Taylor, *Lifting modular mod $l$ representations*, Duke Math. J. **74** (1994), no. 2, 253–269.

[11] M. Harris and R. Taylor, *Regular models of certain Shimura varieties*, Asian J. Math. **6** (2002), no. 1, 61–94.

[12] L. Illusie, *Réalisation $l$-adique de l'accouplement de monodromie d'après A. Grothendieck*, Astérisque **7** (1992), 27–44.

[13] C. Khare, *On isomorphisms between deformation rings and Hecke rings*, Invent. Math. **154** (2003), no. 1, 199–222.

[14] C. Khare and R. Ramakrishna, *Finiteness of Selmer groups and deformation rings*, Invent. Math. **154** (2003), no. 1, 179–198.

[15] B. Mazur, *An introduction to the deformation theory of Galois representations*, in *Modular forms and Fermat's Last Theorem* (Boston, MA, 1995), 243–311, Springer, New York, 1997.

[16] L. Nyssen, *Pseudo-représentations*, Math. Ann. **306** (1996), no. 2, 257–283.

[17] R. Ramakrishna, *On a variation of Mazur's deformation functor*, Compositio Math. **87** (1993), no. 3, 269–286.

[18] K. A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), no. 2, 431–476.

[19] K. A. Ribet and S. Takahashi, *Parametrizations of elliptic curves by Shimura curves and by classical modular curves*, Proc. Nat. Acad. Sci. U.S.A. **94** (1997), no. 21, 11110–11114.

[20] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[21] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[22] L. Yang, *Multiplicities of galois representations in the higher weight sheaf cohomology associated to shimura curves*, Ph.D. thesis, City University of New York, 1996.

Yih-Jeng Yu
Department of Mathematical Sciences, National Chengchi University, Taipei 11605, Taiwan
*E-mail address*: yjyu@nccu.edu.tw