# REGULAR MODELS OF CERTAIN SHIMURA VARIETIES*

## MICHAEL HARRIS[†] AND RICHARD TAYLOR[‡]

**0. Introduction.** This paper develops a technique for studying the bad reduction of Shimura varieties attached to twisted unitary groups. The Shimura varieties we consider are the precise analogues in higher dimension of the modular curves $X_1(p)$. It turns out that the theory is completely analogous to the well-known theory for modular curves, due to Deligne and Rapoport [DR].

Our original purpose in investigating these special cases of bad reduction was with a view to extending the techniques of Taylor-Wiles [TW] to deformations of mod $\ell$ Galois representations of dimension $> 2$. The present article has been extracted from an unpublished manuscript, mostly written between 1996 and 1998, which largely realized this objective, though undoubtedly not in optimal form [HT1]. Subsequent developments, especially an idea of Skinner and Wiles [SW] for bypassing Ribet's level-lowering argument, lead us to hope that more complete results may be within reach.

The Taylor-Wiles method requires information about the tame ramification at $q$ of the Galois representations associated to modular forms of level $\Gamma_1(q)$ for certain primes $q$ highly congruent to 1 modulo $\ell$. The extension of this method to higher dimension is based on the detailed study of the singularities of the special fiber at $q$ for level subgroups generalizing $\Gamma_1(q)$. In the interim, our work on the local Langlands conjecture required a much more comprehensive study of bad reduction of Shimura varieties. The results of [HT2] apply in all levels, and yield the Galois-theoretic statements of the present paper as a special case.

Nevertheless, we feel the present paper is of independent interest. In the first place, the results presented here are somewhat more precise than those of [HT2], in the situations to which they apply. In particular, the congruence formula for the $U$-operator, proved in §4.2, is not stated explicitly as such in [HT2]. (A congruence formula in the analogue of level $\Gamma_0(q)$, valid in considerable generality, is due to T. Wedhorn [We2]; we treat the analogue of level $\Gamma_1(q)$.) In the second place, our method, based on the Tate-Oort classification of group schemes of order $p$, is relatively elementary, and is likely to apply to Shimura varieties whose geometry is not amenable to the sort of analysis carried out in [HT2].

Finally, the modular curves $X_1(p)$ play a special role in the theory of $p$-adic modular forms, as developed by Hida and Coleman, as well as in the related work of Gross [Gr]. We hope the approach to bad reduction developed here will contribute to the extension of this theory to modular forms in higher dimension.

The authors would like to thank the University of Durham (Harris and Taylor), Harvard University (Harris), the Institut Henri Poincaré (Taylor), and RIMS and Kyoto University (Harris) for providing a hospitable working environment at different stages of the project. We are particularly grateful to A. J. de Jong for suggesting we

use the Tate-Oort classification for passing from level $\Gamma_0$ to level $\Gamma_1$, as in [DR]. We also thank Gabber, Rapoport, and Zink for help in correcting sign errors in an early version of the manuscript, and Labesse for numerous conversations regarding base change. Finally, we thank Genestier for pointing out some bothersome notational inconsistencies, and suggesting a remedy.

## 1. Notation for unitary groups.

**1.1.** Let $E$ be a totally real number field of degree $d$ over $\mathbb{Q}$ and let $\mathcal{K}$ be a totally imaginary quadratic extension of $E$; let $c \in Gal(\mathcal{K}/E)$ denote the non-trivial automorphism. We assume $\mathcal{K}$ contains an imaginary quadratic field $\mathcal{K}_0$, so that $\mathcal{K} = \mathcal{K}_0 \cdot E$. We denote by $\Sigma_E$ and $\Sigma_\mathcal{K}$ the sets of complex embeddings of $E$ and $\mathcal{K}$. Let $D$ be a central simple algebra of dimension $n^2$ over $\mathcal{K}$, endowed with an involution, denoted $\tilde{c}$, that induces the Galois automorphism $c$ on $\mathcal{K}$; i.e., $\tilde{c}$ is an involution of the second kind.

We define algebraic groups $U(D) = U(D, \tilde{c})$ and $GU(D) = GU(D, \tilde{c})$ over $\mathbb{Q}$ such that, for any $\mathbb{Q}$-algebra $R$,

$$U(D)(R) = \{g \in D^{opp} \otimes_\mathbb{Q} R \mid g \cdot \tilde{c}(g) = 1\};$$

$$GU(D)(R) = \{g \in D^{opp} \otimes_\mathbb{Q} R \mid g \cdot \tilde{c}(g) = \nu(g) \text{ for some } \nu(g) \in R^\times\}.$$

Thus $GU(D)$ admits a homomorphism $\nu : GU(D) \to \mathbb{G}_m$ with kernel $U(D)$. There is an algebraic group $U_E(D)$ over $E$ such that $U(D) \xrightarrow{\sim} R_{E/\mathbb{Q}} U_E(D)$, where $R_{E/\mathbb{Q}}$ denotes Weil's restriction of scalars functor. This isomorphism identifies automorphic representations of $U(D)$ and $U_E(D)$.

The groups $U(D)$ (resp. $GU(D)$) are all inner forms of the same quasi-split unitary group (resp. unitary similitude group), denoted $U_0$ (resp. $GU_0$). The group $U_0$ is of the form $U(D_0, \tilde{\chi}(*)_0)$ where $D_0$ is the matrix algebra and $\tilde{\chi}(*)_0$ is an appropriate involution. Then $U_{0,\infty} \cong U(\frac{n}{2}, \frac{n}{2})^{[E:\mathbb{Q}]}$ if $n$ is even, $U_{0,\infty} \cong U(\frac{n+1}{2}, \frac{n-1}{2})^{[E:\mathbb{Q}]}$ if $n$ is odd.

Let $G$ be a reductive algebraic group over the number field $E$. If $v$ is a place of $E$ we let $G_v = G(\mathbb{Q}_v)$; if $v$ is archimedean we let $\mathfrak{g}_v = Lie(G_v)_\mathbb{C}$. We let $G_\infty$ denote $\prod_{v|\infty} G_v$, the product taken over all archimedean places of $E$, and let $\mathfrak{g}_\infty = \prod_{v|\infty} \mathfrak{g}_v$.

Let $\pi$ be an irreducible automorphic representation of $G$; i.e., an irreducible $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbf{A}^f)$-module that embeds as a submodule of the space of automorphic forms relative to the chosen maximal compact subgroup $K_\infty$. We write $\pi = \pi_\infty \otimes \pi_f$ as usual, and say $\pi$ is *cohomological* if $\pi$ is cuspidal and if the relative Lie algebra cohomology $H^\bullet(\mathfrak{g}_\infty, K_\infty; \pi_\infty \otimes V) \neq 0$ for some finite dimensional representation $V$ of $\mathfrak{g}_\infty$. We let $Coh(G)$ denote the set of cohomological cuspidal automorphic representations of $G$, $Coh(G, V) \subset Coh(G)$ the subset of $\pi$ for which $H^\bullet(\mathfrak{g}_\infty, K_\infty; \pi_\infty \otimes V) \neq 0$, with $V$ fixed. If $K \subset G(\mathbf{A}^f)$ is a compact open subgroup, let $Coh(G, K)$ denote the set of $\pi \in Coh(G)$ such that $\pi^K \neq \{0\}$, $Coh(G, K, V) = Coh(G, K) \cap Coh(G, V)$.

Let $\mathcal{A}_0(G)$ denote the space of cusp forms on $G$. Let $Rep(G)$ denote the set of equivalence classes of irreducible $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbf{A}^f)$-modules. If $\pi \in Rep(G)$, we let $m(\pi) = \dim Hom(\pi, \mathcal{A}_0(G))$, where $Hom$ denotes the space of homomorphisms of $(\mathfrak{g}_\infty, K_\infty) \times G(\mathbf{A}^f)$-modules. More generally, let $S$ be a finite set of places of $E$, containing the archimedean places (for simplicity) and let $Rep(G)^S$ denote the set of equivalence classes of irreducible $G(\mathbf{A}^{f,S})$-modules, where $G(\mathbf{A}^{f,S}) \subset G(\mathbf{A}^f)$ is the subgroup with trivial entry at every place in $S$. We say $\pi^S \in Rep(G)^S$ is *automorphic*

if

$$Hom_{G(\mathbf{A}^f,s)}(\pi^S, \mathcal{A}_0(G)) \neq 0;$$

i.e., if $\pi^S$ can be extended to a cuspidal automorphic representation of $G$. We say $\pi^S$ is *cohomological* if it can be extended to a cohomological cuspidal automorphic representation of $G$.

We now fix a rational prime $p$ that splits in $\mathcal{K}_0$ and an archimedean prime $s_0$ of $E$. We choose a central simple algebra $D^{\#}$ over $\mathcal{K}$, with involution of the second kind $\tilde{c}^{\#}$. We write $G = GU(D^{\#}, \tilde{c}^{\#})$, $G' = U(D^{\#}, \tilde{c}^{\#})$, and view $G'$ alternatively as a group over $\mathbb{Q}$ or $E$. We assume that

$$(1.1.1) \qquad G'_{\infty} \cong U(n-1, 1) \times U(n)^{[E:\mathbb{Q}]-1}.$$

Moreover, if $v$ is a finite prime, we assume that

$$(1.1.2) \qquad G'_v \cong U_{0,v} \text{ if } v \text{ does not split in } \mathcal{K}/E.$$

In the applications, we will always assume $D^{\#}$ to be a division algebra.

We need another involution $\tilde{c}^*$ of the second kind on $D^{\#}$; $\tilde{c}^*$ is a *positive* involution, so that the unitary group $U(D, \tilde{c}^*)$ is totally definite at infinity. The existence of such involutions is established as in the introduction to [HT2].

**1.2. Open compact subgroups.** Fix an open compact subgroup $K = \prod_v K_v \subset G(\mathbf{A}^f)$, and define

$$(1.2.1) \qquad \mathfrak{S}_K(\mathbb{C}) = G(\mathbb{Q}) \backslash G(\mathbf{A}) / Z_G(\mathbb{R}) \cdot K_{\infty} \cdot K; \quad \mathfrak{S}(\mathbb{C}) = \varprojlim \mathfrak{S}_K(\mathbb{C});$$

the limit taken with respect to inclusion. Then $\mathfrak{S}_K$ is the locally symmetric space associated to $G$ and its subgroup $K$. The structure of the projective limit $\mathfrak{S}(\mathbb{C})$ as the set of complex points of a Shimura variety will be recalled below.

We will be working with specific choices of local subgroups $K_v$. For any rational prime $q$ that splits in $\mathcal{K}_0$, we choose a maximal compact subgroup of $G_q$ in the form

$$(1.2.2) \qquad \mathbb{Z}_q^{\times} \times \prod_{v \mid q} G_v,$$

the product being taken over divisors $v$ of $q$ in $E$. Here if $w$ is a divisor of $q$ in $\mathcal{K}_0$ lying above $v$ and if $D_w$ is isomorphic to $GL(a, B_w)$ for some factorization $n = ab$ and some division algebra $B_w$ of degree $b^2$ over $\mathcal{K}_w$, then $G_v$ can be taken in the form $GL(a, O_{B_w})$, where $O_{B_w}$ is the maximal order of $B_w$.

We choose a finite set $Q$ of finite places of $E$, each dividing a distinct rational prime that splits in $\mathcal{K}_0$, and an additional finite place $\{\mathfrak{r}\}$ of $E$, also split in $\mathcal{K}$. Primes $\mathfrak{q} \in Q$ are assumed to have the property that, if $q$ is the rational prime divisible by $\mathfrak{q}$, then $q$ splits completely in $E$. We let $Q(\mathbb{Q})$ denote the set of rational primes divisible by primes in $Q$. For $\mathfrak{q} \in Q$ we define

$$\Gamma_{0,\mathfrak{q}} = \{k \in GL(n, \mathbb{Z}_q) \mid k \equiv \begin{pmatrix} *_1 & * \\ 0 & *_{n-1} \end{pmatrix} \pmod{q}\}$$

$$\Gamma_{1,\mathfrak{q}} = \{k \in GL(n, \mathbb{Z}_q) \mid k \equiv \begin{pmatrix} 1 & * \\ 0 & *_{n-1} \end{pmatrix} \pmod{q}\}$$

as subgroups of the $\mathfrak{q}$-factor of 1.2.2. Then we let

$$K_{0,q} = \mathbb{Z}_q^\times \times \prod_{v \neq \mathfrak{q}, v | q} GL(n, \mathcal{O}_v) \times \Gamma_{0,\mathfrak{q}};$$

$$K_{1,q} = \mathbb{Z}_q^\times \times \prod_{v \neq \mathfrak{q}, v | q} GL(n, \mathcal{O}_v) \times \Gamma_{1,\mathfrak{q}},$$

viewed as subgroups of $G_q$.

Let $q(\mathfrak{r})$ denote the residue characteristic of $\mathfrak{r}$. Let $N_0$ denote the upper triangular unipotent subgroup of $GL(n)$ and let

$$I_1(\mathfrak{r}) = \{k \cdot n \in GL(n, \mathcal{O}_\mathfrak{r}) \mid k \equiv 1 \pmod{\mathfrak{r}}, n \in N_0(\mathcal{O}_\mathfrak{r})\};$$

$$I(\mathfrak{r}) = \mathbb{Z}_{q(\mathfrak{r})}^\times \times \prod_{v \neq \mathfrak{r}, v | q(\mathfrak{r})} GL(n, \mathcal{O}_w) \times I_1(\mathfrak{r}).$$

We let

(1.2.3)                                $K_{q(\mathfrak{r})} = I_\mathfrak{r}.$

LEMMA 1.3. *For $q(\mathfrak{r})$ sufficiently large, the locally symmetric variety $\mathfrak{S}_K(\mathbb{C})$ is smooth for $K = K_{0,Q}$ or $K_{1,Q}$. Moreover, for any $s \in G(\mathbf{A}^f)$ the groups $s^{-1}G(\mathbb{Q})s \cap K_{0,Q}$ and $s^{-1}G(\mathbb{Q})s \cap K_{1,Q}$ are trivial.*

*Proof.* The first assertion follows from the second. Let $K = K_{0,Q}$ or $K_{1,Q}$. Let $x \in s^{-1}G(\mathbb{Q})s \cap K$ for some $s \in G(\mathbf{A}^f)$. The subgroup of $G(\mathbf{A})$ generated by $x$ is both discrete and compact, hence finite. The group $I_\mathfrak{r}$ is pro-$q(\mathfrak{r})$ and it follows that $x$ is a root of unity of order a power of $q(\mathfrak{r})$, lying in some extension field $\mathcal{K}'$ of $\mathcal{K}$ that admits an embedding in $D^\#$. The degree of $\mathcal{K}'$ over $\mathbb{Q}$ is bounded by $n[\mathcal{K} : \mathbb{Q}]$, hence for $q(\mathfrak{r})$ sufficiently large we must have $x = 1$.

**1.4. Cohomology.** Let $V$ be a finite-dimensional irreducible representation of $\mathfrak{g}_\infty$, and let $\pi$ be an automorphic representation in $Coh(G, K, V)$. Let $\mathcal{E}$ be a field of definition for $V$, finite over $\mathbb{Q}$. Fix a rational prime $\ell$ and an embedding of $\mathcal{E}$ in $\overline{\mathbb{Q}}_\ell$, and let $V(\overline{\mathbb{Q}}_\ell)$ be the $\overline{\mathbb{Q}}_\ell$-form of $V$ defined by extension of scalars from $\mathcal{E}$. Let

(1.4.1)            $\tilde{V}_\ell = G(\mathbb{Q}) \backslash G(\mathbf{A}) \times V(\overline{\mathbb{Q}}_\ell) / Z_G(\mathbb{R}) \cdot K_\infty \cdot K$

be the $\ell$-adic local system over $\mathfrak{S}_K$ associated to $V$ and our chosen embedding of $\mathcal{E}$ in $\overline{\mathbb{Q}}_\ell$. The definition depends on $K$ but the local systems for $K' \subset K$ are compatible with the natural map $\mathfrak{S}_{K'} \to \mathfrak{S}_K$, so we omit $K$ from the notation for $\tilde{V}_\ell$. We define

$$H^\bullet(\mathfrak{S}(\mathbb{C}), \tilde{V}_\ell) = \varinjlim_K H^\bullet(\mathfrak{S}_K(\mathbb{C}), \tilde{V}_\ell).$$

Since $\pi_f$ occurs in the cohomology of the complex local system associated to $V$, it admits a form, also denoted $\pi_f$ over a number field $E(\pi)$. Let $\lambda$ be a prime of $E(\pi)$ dividing $\ell$, and let $\pi_{f,\lambda} = \pi_f \otimes_{E(\pi)} E(\pi)_\lambda$; we apply the same notation to the local components of $\pi_f$. Let

(1.4.2)        $M_\lambda^\bullet[\pi_f] = H^\bullet(\mathfrak{S}(\mathbb{C}), \tilde{V}_\ell)[\pi_f] = Hom_{G(\mathbf{A}^f)}(\pi_{f,\lambda}, H^\bullet(\mathfrak{S}(\mathbb{C}), \tilde{V}_\ell)),$

a finite-dimensional vector space over $E(\pi)_\lambda$.

Let $\pi$ be an automorphic representation of $G$. The restriction of $\pi$ to the unitary group $G'$, which can be viewed as an algebraic group over $E$, decomposes as a direct sum of irreducible automorphic representations. Any two summands have the same local components at any finite place $w$ dividing a rational prime $q$ that splits in $\mathcal{K}_0$, since at such places the similitude map splits as a product $G(\mathbb{Q}_q) \simeq G'(\mathbb{Q}_q) \times \mathbb{Q}_q^\times$. For such a place $w$, we say $\pi$ is **non-monodromic at w** if one component (hence every component) of the restriction of $\pi$ to the unitary group $G'$ corresponds under the local Langlands correspondence to a semisimple representation of the Weil-Deligne group of $E_w$ (if $G_w$ is split) or corresponds to a supercuspidal representation of $GL(n, E_w)$ by the Jacquet-Langlands correspondence (if $G_w$ is the multiplicative group of a division algebra).

PROPOSITION 1.4.3. *Let $\pi$ be an automorphic representation of $G$ such that $\pi^{K_{1,q}} \neq 0$. Suppose $\pi_f$ is cohomological, and occurs only in cohomology in the middle degree $n-1$. Then $\pi_q$ is generic for every $q$ that splits in $\mathcal{K}_0$. For every $w \in Q$, $\pi_w$ is either (a) unramified; (b) principal series attached to an n-tuple $(\alpha, \beta_1, \ldots, \beta_{n-1})$ of characters of $E_w^\times$, with $\alpha$ tamely ramified and each $\beta_i$ unramified; or (c) the Langlands sum of a special representation of $GL(2)$ and an unramified representation of $GL(n-2)$. In cases (a) and (c), but not in case (b), $\pi_w$ has a $\Gamma_{0,w}$-fixed vector.*

*Proof.* The first assertion follows from the existence and properties of base change of $\pi$ to $G_{\mathcal{K}_0} \simeq GL(n)_{\mathcal{K}} \times GL(1)_{\mathcal{K}_0}$, [CL, L, cf. HT2, VI.2]. Let $\Pi$ denote the base change. Since $\pi_f$ occurs only in the middle degree, it follows from the results of [C2] that $\Pi$ is cuspidal, hence globally generic by Shalika's theorem. The genericity of $\pi_q$ is then a consequence of Theorem 4.6.2 of [L]. The remaining assertions then follow easily from the Bernstein-Zelevinsky classification of admissible irreducible representations of $GL(n)$ and from the theory of the conductor [JPSS].

REMARK 1.4.4 The hypothesis that $\pi_f$ occurs only in the middle degree is standard, and is automatically satisfied, for instance, if $\pi$ is obtained by descent [C1,CL] from a cuspidal automorphic cohomological representation of $GL(n)_{\mathcal{K}}$, or if $\pi_f$ is supercuspidal at one place that splits in $\mathcal{K}$ (cf. [H1]).

**1.5. Notation for unramified Hecke algebras.** Let $\mathcal{CS}(\mathbb{Q})$ denote the set of all primes of $\mathbb{Q}$, and let $\mathcal{CS}^+(\mathbb{Q})$ be the subset of finite primes that split in $\mathcal{K}_0$ and are unramified in $\mathcal{K}$. For $v \in \mathcal{CS}^+(\mathbb{Q})$ of residue characteristic $p$ we choose a place $v_1$ of $\mathcal{K}_0$ above $v$, and let $\Sigma = \Sigma(v_1)$ be the set of primes of $\mathcal{K}$ dividing $v_1$. Then we have

$$(1.5.1) \qquad\qquad G_v \cong \prod_{w \in \Sigma} GL(n, \mathcal{K}_w) \times \mathbb{Q}_v^\times .$$

The Hecke algebra $\mathbf{T}_v$ of $G_v$ relative to any maximal compact subgroup (conjugate to $\prod_{w \in \Sigma} GL(n, \mathcal{O}_{\mathcal{K}_w}) \times \mathbb{Z}_v^\times$) is isomorphic to a polynomial algebra over $\mathbb{Z}[\frac{1}{p}]$ in the variables

$$\{T_{i,w}, i = 1, \ldots, n, T_{n,w}^{-1}, w \in \Sigma; T_{0,v}, T_{0,v}^{-1}\}.$$

Here $\mathbb{Z}[T_{0,v}, T_{0,v}^{-1}]$ is the Hecke algebra of the factor $\mathbb{Q}_v^\times$ in 1.5.1; by abuse of language we refer to the $T_{i,w}$ as the Hecke operators at $w$, or at the prime of $E$ below $w$. The

Hecke operators at $w$ are normalized so that

$$(1.5.2) \qquad\qquad P_w(q^{-s}) = 1 + \sum_{i=1}^{n} (-1)^i T_{i,w} q^{-is}$$

is the local Euler factor at $w$ of the motivically normalized standard $L$-function of $GL(n)$. Here $q$ is the order of the residue field $k(w)$ and the inverse roots of $P_w(X)$ are the Satake parameters, multiplied by $q^{(n-1)/2}$. Up to canonical isomorphism the algebra $\mathbf{T}_v$ does not depend on the choice of $v_1$ above $v$.

The global Hecke algebra $\mathbf{T}$ is the tensor product over $v \in \mathcal{CS}^+(\mathbb{Q})$ of the $\mathbf{T}_v$. If $S$ is a finite subset of $\mathcal{CS}^+(\mathbb{Q})$, we let $\mathbf{T}^S \subset \mathbf{T}$ be the subalgebra generated by the $\mathbf{T}_v$ for $v \notin S$.

## 2. Shimura varieties and Galois representations.

**2.1. Shimura varieties defined by twisted and untwisted unitary groups.** Let $E$ be a totally real field of degree $d$, $\mathcal{K}_0$ an imaginary quadratic field, $\mathcal{K} = \mathcal{K}_0 \cdot E$. As in 1.1, we let $G$ be the similitude group of a division algebra $D^{\#}$ of dimension $n^2$ over $\mathcal{K}$ with involution $\tilde{c}^{\#}$ of the second kind (with rational similitude factor). We will always assume $n > 2$. The unitary group $G' \subset G$ is assumed to satisfy (1.1.1) at real places and (1.1.2) at finite places; we let $s_1$ denote the real place of $E$ at which $G'$ has signature $(n-1, 1)$. The field $\mathcal{K}_0$ is assumed to be given with a fixed complex embedding $1_{\mathcal{K}_0}$, and we let $\Phi$ be the set of complex embeddings of $\mathcal{K}$ inducing $1_{\mathcal{K}_0}$ on $\mathcal{K}_0$. Then $\Phi$ is a CM type of $\mathcal{K}$. Let $t_1 \in \Phi$ be the complex place above $s_1$.

We use $\Phi$ to identify $D^{\#} \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow{\sim} M(n, \mathbb{C})^d$. In terms of this identification, we define

$$h_\Phi : \mathbb{C}^\times \simeq R_{\mathbb{C}/\mathbb{R}} \mathbb{G}_{m,\mathbb{C}} \to (D^{\#,opp} \otimes_{\mathbb{Q}} \mathbb{R})^\times$$

by

$$(2.1.1) \qquad\qquad h_\Phi(z) = \left( \begin{pmatrix} z \cdot I_{n-1} & \\ & \bar{z} \end{pmatrix}, z I_n, \ldots, z I_n \right),$$

where the first matrix corresponds to $t_1$. We may view $h_\Phi$ as a homomorphism with values in $G_{\mathbb{R}}$. Let $X = X_{n-1}$ be the $G_{\mathbb{R}}$-conjugacy class of homomorphisms from $R_{\mathbb{C}/\mathbb{R}} \mathbb{G}_{m,\mathbb{C}}$ to $G_{\mathbb{R}}$ containing $h_\Phi$. Then $X$ has a natural $G_{\mathbb{R}}$-invariant complex structure which is isomorphic to the unit ball in $\mathbb{C}^{n-1}$. The pair $(G, X)$ is the datum defining the Shimura variety $\mathfrak{S}$ introduced in (1.2.1), and considered in [C2,K,H1]. As in [H1], we see that the reflex field (Shimura field) $E(G, X)$ is just $t_1(\mathcal{K})$.

**2.1.2.** We fix a maximal order $\mathcal{O}^{\#} \subset D^{\#}$, stable under $\tilde{c}^*$, and let $\hat{\mathcal{O}}^{\#} = \mathcal{O}^{\#} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$. Let $V = D^{\#}$, viewed as a $2dn^2$-dimensional vector space over $\mathbb{Q}$ with an embedding $D^{\#} \to End(V)$. The homomorphism $h_\Phi$ takes values in $D^{\#,opp} = End_{D^{\#}}(V)$, and is defined over some extension $L$ of $\mathcal{K}$ which splits $D^{\#}$. There is then an $L$-rational decomposition

$$(2.1.2.1) \qquad\qquad V \otimes_{\mathbb{Q}} L \xrightarrow{\sim} V_0 \oplus V_1,$$

where $V_0$ (resp. $V_1$) is the $\bar{z}$- (resp. $z$-)eigenspace for $h(z)$; the subspaces $V_0$ and $V_1$ are stable under the left action of $D^{\#} \otimes_{\mathbb{Q}} L$.

We restrict our attention to open compact subgroups $K \subset \hat{\mathcal{O}}^{\#,\times}$, though this is only for convenience. Moreover, we always assume $K$ factors as the product of subgroups of $G(\mathbb{Q}_p)$, as $p$ varies over rational primes. We consider the moduli problem $\mathcal{A}_K$, defined on the category of schemes over $Spec\, E(G,X)$. For $T$ a scheme over $Spec\, E(G,X) = Spec\, \mathcal{K}$, the functor $\mathcal{A}_K(T)$ consists of equivalence classes (see below) of quadruples $(A, \iota, \lambda, \beta)$ consisting of

**(2.1.2.2)** An abelian scheme $A$ over $T$ of dimension $dn^2$;

**(2.1.2.3)** An embedding $\iota : \mathcal{O}^\# \to End_T(A)$;

**(2.1.2.4)** A polarization $\lambda : A \xrightarrow{\sim} \hat{A}$, such that

$$\lambda \circ \iota(b) = \iota(\widehat{c^*(b)}) \circ \lambda$$

for all $b \in \mathcal{O}^\#$;

**(2.1.2.5)** An $\hat{\mathcal{O}}^\#$-equivariant level $K$ structure

$$\beta : H_1(A, \hat{\mathbb{Z}}) = \prod_q T_q(A) \xrightarrow{\sim} \hat{\mathcal{O}}^\# \quad (\text{mod } K),$$

in the sense of Kottwitz ([K2], pp. 390-391; cf. [HT2]), compatible with $\lambda$.

These data are assumed to satisfy the determinant condition of Kottwitz, namely

$$(2.1.2.6) \qquad det_{\mathcal{O}_T}(b; Lie A) = det_L(b; V_1), \qquad b \in \mathcal{O}^\#.$$

We consider two quadruples $(A, \iota, \lambda, \beta)$ and $(A', \iota', \lambda', \beta')$ equivalent if there is an isogeny $\alpha : A \to A'$ which takes $(\lambda, \iota, \beta)$ to $(t\lambda', \iota', \beta')$, where $t$ is a positive rational number.

For $K$ sufficiently small the functor $\mathcal{A}_K$ is representable by a smooth projective variety, also denoted $\mathcal{A}_K$, over $E(G,X)$ [K2,§5]. For general $K$ we take a normal subgroup of finite index $K' \subset K$ for which $\mathcal{A}_{K'}$ is representable and then let $\mathcal{A}_K$ be the scheme-theoretic quotient of $\mathcal{A}_{K'}$ by $K/K'$. Then the Shimura variety $\mathfrak{S}_K$, viewed as a scheme over $Spec\, E(G,X)$, is an open and closed subscheme of $\mathcal{A}_K$. More precisely, $\mathcal{A}_K$ is isomorphic to a finite union of Shimura varieties of the form $\mathfrak{S}_K$, the number of Shimura varieties involved being given by the deviation from the Hasse principle for $G$ (cf. [K2,§8; RZ2, pp. 301-302]). In particular, $\mathfrak{S}_K$ is projective for all $K$, and is smooth provided $K \cap (\mathcal{O}^\#)^\times = \{1\}$.

**Warning 2.1.2.7** Note that the level structure $\beta$ defined here goes in the direction opposite to the level structures $\eta$ introduced in [HT2,IV.1], as well as in [K2]. This convention, which is justified on Hodge-theoretic grounds, leads to a dualization in the associated Galois representations.

**2.1.3.** The level structures corresponding to the open compact subgroups of the form $K_{0,Q}$ and $K_{1,Q}$, defined as in 1.2, deserve special attention. Recall that $Q$ denotes a finite set of primes of $E$ dividing rational primes $q$ that split completely in $\mathcal{K}$, and such that $D^\#$ is split at all primes dividing $q$. In particular, we may take $L = \mathbb{Q}_q$ in (2.1.2.1); then

$$(2.1.3.1) \qquad V \otimes_{\mathbb{Q}} L \cong \bigoplus_{w \mid q} M(n, \mathcal{K}_w) \cong \bigoplus_{w \mid q} M(n, \mathbb{Q}_q)^2,$$

where $w$ runs through primes of $E$ (note: *not* $\mathcal{K}$) dividing $q$ and the exponent 2 corresponds to the quadratic extension $\mathcal{K}/E$. We let $V_w = M(n, \mathcal{K}_w)$ in the above

decomposition. Recall that we are identifying $\mathcal{K}$ with the subfield $E(G, X)$ of $\mathbb{C}$ via $t_1$. Then the first coordinate of $h_\Phi$, in the representation (2.1.1), picks out one summand of (2.1.3.1), say the summand $V_{\mathfrak{q}}$ corresponding to the prime $w = \mathfrak{q}$. Write $V_{\mathfrak{q}} = V_{\mathfrak{q}^{(1)}} \oplus V_{\mathfrak{q}^{(2)}}$, where $\mathfrak{q}^{(1)}$ and $\mathfrak{q}^{(2)}$ are the primes of $\mathcal{K}$ above $w$. These can be numbered so that

$$(2.1.3.2) \quad \dim_{\mathbb{Q}_q} V_0 \cap V_{\mathfrak{q}^{(1)}} = n; \quad \dim_{\mathbb{Q}_q} V_1 \cap V_{\mathfrak{q}^{(1)}} = n(n-1)$$
$$\dim_{\mathbb{Q}_q} V_0 \cap V_{\mathfrak{q}^{(2)}} = n(n-1); \quad \dim_{\mathbb{Q}_q} V_1 \cap V_{\mathfrak{q}^{(2)}} = n.$$

The corresponding maximal ideals of $\mathcal{O}_{\mathcal{K}}$ are denoted $\mathfrak{m}_{\mathfrak{q}^{(i)}}$, $i = 1, 2$.

We let the $\mathfrak{q}$ chosen in the preceding paragraph be the prime in $Q$ dividing $q$, and we define $K_{0,q}$ and $K_{1,q}$ as in 1.2 with respect to this choice of $w = \mathfrak{q}$. Thus $K_{?,Q} = K_{?,q} \times K^q$, where $? \in \{0, 1\}$ and $K^q$ denotes the product of the level subgroups at primes away from $q$. Then the datum $\beta$ breaks up into $q$-primary and prime-to-$q$ components, denoted in the obvious notation $\beta_q$ and

$$(2.1.3.3) \qquad \beta^q : \prod_{p \neq q} T_p(A) \xrightarrow{\sim} (\hat{O}^\#)^q \quad (\mathrm{mod}\ K^q),$$

respectively. First consider $K_{0,Q}$. We let

$$\Gamma_q = GU(\mathcal{O}^\# \otimes_{\mathbb{Z}} \mathbb{Z}_q)^\times \simeq \prod_{w | q} GL(n, \mathbb{Z}_q) \times \mathbb{Z}_q^\times,$$

so that $\Gamma_q / K_{0,q}$ is naturally in bijection with $\mathbb{P}^{n-1}(k(\mathfrak{q}))$, where $k(\mathfrak{q}) \simeq \mathbb{F}_q$ is the residue field of $\mathfrak{q}$. Let $K^+ = \Gamma_q \times K^q$.

Let $\varpi$ be a uniformizer in $E_{\mathfrak{q}}$. We write $t(\varpi)_1$ for the diagonal matrix with entry $\varpi^{-1}$ as first entry and 1 elsewhere, viewed as an element of $G(\mathbb{Q}_q) \subset G(\mathbf{A}_f)$. Then

$$(2.1.3.4) \qquad K_{0,Q} = K^+ \cap t(\varpi)_1^{-1} K^+ t(\varpi)_1.$$

We thus have two maps $f_i : \mathfrak{S}_{K_{0,Q}} \to \mathfrak{S}_{K^+}$, where $f_1$ is defined by the inclusion $K_{0,Q} \subset K^+$ and $f_2$ is defined by inclusion in $t(\varpi)_1^{-1} K^+ t(\varpi)_1$, followed by right-multiplication by $t(\varpi)_1^{-1}$. Let $x = (A, \iota, \lambda, \beta)$ be a geometric point of $\mathfrak{S}_K$, and let $f_2(x) = (A', \iota', \lambda', \beta')$. Then $x$ is determined by the quintuple $(A, A', \iota, \lambda, \beta^q)$.

By modifying our definition of $f_2$, we can arrange to have an isogeny $A \to A'$, and then realize $\mathfrak{S}_{K_{0,Q}}$ as (an open and closed subscheme of) the moduli space parametrizing quintuples as above, with $A \to A'$ an isogeny of a certain type. However, we do not need an actual isogeny. Let $A$ be an abelian scheme over $T$ as in (2.1.2.2), and let $\mathbf{X}_q(A)$ be the associated $q$-divisible group. Then $\mathbf{X}_q(A)$ inherits an action of $\mathcal{O}^\# \otimes_{\mathbb{Z}} \mathbb{Z}_q$. Let $\mathfrak{q}^{(2)}$ be the prime of $\mathcal{K}$ chosen above. Let $e_2$ be an elementary idempotent in $M(n, \mathbb{Z}_q)$, which we identify with $\mathcal{O}_{\mathfrak{q}^{(2)}}^\# = \hat{O}^\# \cap D_{\mathfrak{q}^{(2)}}^\#$. Then $e_2 \cdot \mathbf{X}_q(A)$ is a $q$-divisible group of height $n$, which we denote $\mathbf{X}(A)_2$. With our choice of $\mathfrak{q}^{(2)}$, it follows easily from (2.1.2.6) and (2.1.3.2) that $\mathbf{X}(A)_2$ is of dimension 1. Now the rational Tate modules $V_q(A) = T_q(A) \otimes \mathbb{Q}_q$ and $V_q(A')$ can be identified. One sees easily that $e_2 \cdot T_q(A') \supset e_2 \cdot T_q(A)$; hence there is an isogeny

$$(2.1.3.5) \qquad \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$$

with kernel $\mathcal{O}_{\mathcal{K}}$ isomorphic to $\mathcal{O}_{\mathcal{K}} / \mathfrak{q}^{(2)} \cong \mathbb{Z} / q\mathbb{Z}$. We conclude that

**2.1.3.6.** For $K^q$ sufficiently small, the scheme $\mathfrak{S}_{K_0,Q}$ is an open and closed subscheme of the moduli scheme parametrizing quintuples

$$(A, \iota, \lambda, \beta^q, \mathbf{X}(A)_2 \to \mathbf{X}(A')_2),$$

where $(A, \iota, \lambda)$ are as in (2.1.2.2-4), $\beta^q$ is a level $K^q$-structure as in (2.1.3.3), and $\mathbf{X}(A)_2 \to \mathbf{X}(A')_2$ is a $q$-isogeny of $q$-divisible $\mathcal{O}_{\mathfrak{q}^{(2)}}$-modules of height $n$ and dimension 1, the whole assumed to satisfy the determinant condition (2.1.2.6).

Here is an alternative definition of the moduli problem. Define $\mathbf{X}(A)_1 = e_1 \cdot \mathbf{X}_q(A)$ by analogy with $\mathbf{X}(A)_2$ above, where $e_1$ is an elementary idempotent in $\mathcal{O}^{\#}_{\mathfrak{q}^{(1)}}$. We choose the pair $(e_1, e_2)$ in such a way that the polarization $\lambda$ places $\mathbf{X}(A)_1$ and $\mathbf{X}(A)_2$ in Cartier duality. We consider diagrams of the form

(2.1.4.1)
$$
\begin{array}{ccc}
 & A'' & \\
\phi \swarrow & & \searrow \phi' \\
A & & A'
\end{array}
$$

where $(A, \iota, \lambda, \beta^q)$, $(A', \iota', \lambda', (\beta')^q)$ are as above. In particular $\lambda : A \xrightarrow{\sim} \hat{A}$, $\lambda' : A' \xrightarrow{\sim} \hat{A}'$, are principal polarizations. We assume $\phi$ and $\phi'$ are isogenies of degree $q^n$, with the following property. Let $\hat{\phi} : \hat{A} \to \hat{A}''$ and $\hat{\phi}' : \hat{A}' \to \hat{A}''$ be the dual isogenies to $\phi$, $\phi'$, respectively. Then $Ker(\hat{\phi}')$ (resp. $Ker(\hat{\phi})$) is generated over $\mathcal{O}^{\#}_{\mathfrak{q}^{(2)}}$ (resp. $\mathcal{O}^{\#}_{\mathfrak{q}^{(1)}}$) by a subgroup $\hat{C}_2 \subset \mathbf{X}(A')_2$ (resp. $\hat{C}_1 \subset \mathbf{X}(A)_1$) of order $q$. We let $C_1, C_2 \subset \mathbf{X}(A'')$ be the Cartier duals of $\hat{C}_1$ and $\hat{C}_2$, respectively. Then $\phi$ (resp. $\phi'$) induces an isomorphism $\phi_2 : \mathbf{X}(A'')_2 \xrightarrow{\sim} \mathbf{X}(A)_2$ (resp. $\phi_1 : \mathbf{X}(A'')_1 \xrightarrow{\sim} \mathbf{X}(A')_1$) so that $\phi' \circ (\phi_2)^{-1}$ (resp. $\phi \circ (\phi_1)^{-1}$) defines a $q$-isogeny $r : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$ (resp. $r' : \mathbf{X}(A')_1 \to \mathbf{X}(A)_1$) with kernel canonically isomorphic to $C_2$ (resp. $C_1$).

It is assumed that $\phi$ and $\phi'$ are compatible with the $\mathcal{O}^{\#}$-structure, in the sense that the induced isomorphisms of rational $q$-adic Tate modules are assumed to be $\mathcal{O}^{\#}$-linear. Finally, the two isogenies are assumed to be dual to each other, in the sense that

(2.1.4.2)
$$\hat{\phi}' \circ \lambda' \circ \phi' = \hat{\phi} \circ \lambda \circ \phi : A'' \to \hat{A}''.$$

Using these conditions one reconstructs $A''$ and $(A', \iota', \lambda', (\beta')^q)$ from $(A, \iota, \lambda, \beta^q)$ and $r : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$ in such a way that $\iota'$ and $(\beta')^q$ correspond to $\iota$ and $\beta^q$ via the isogenies. Indeed, we can define $A''$ to be the quotient of $A'$ by the $\mathcal{O}^{\#}_{\mathfrak{q}^{(2)}}$-module generated by $r(\mathbf{X}(A)_2[q])$. The map $\phi'$ is defined so that $A' \to A'' \xrightarrow{\phi'} A'$ is multiplication by $q$, viewed as an element of $\mathcal{O}^{\#}_{\mathfrak{q}^{(2)}}$. Then $\ker(\phi)$ is defined to be $\ker(\hat{\phi}' \circ \lambda' \circ \phi') \cap A''[\mathfrak{q}^{(1)}]$.

It follows easily from (2.1.4.2) that

(2.1.4.3)
$$\ker(\phi') = \widehat{\ker(\phi)} \quad \text{(Cartier duality)}.$$

Indeed, letting $f : A'' \to A''$ denote the map in (2.1.4.2), we see that

$$
\begin{aligned}
e_1 \cdot \ker(f) \times e_2 \cdot \ker(f) &= \ker(f) \cap \mathbf{X}(A'')_1 \times \ker(f) \cap \mathbf{X}(A'')_2 \\
&= e_1 \cdot \ker(\phi) \times e_2 \cdot \ker(\phi') \\
&= e_1 \cdot (\lambda \circ \phi)^{-1} \ker(\hat{\phi}') \times e_2 \cdot (\lambda' \circ \phi')^{-1} \ker(\hat{\phi})
\end{aligned}
$$

This implies $\ker(\hat{\phi}) \xrightarrow{\sim} \ker(\phi')$, which is equivalent to (2.1.4.3). In particular,

$$(2.1.4.4) \qquad C_2 \xrightarrow{\sim} \hat{C}_1 \xrightarrow{\sim} e_2 \cdot \ker(\phi), \quad C_1 \xrightarrow{\sim} \hat{C}_2 \xrightarrow{\sim} e_1 \cdot \ker(\phi'),$$

where $C_1$ and $C_2$ are defined as above.

When $K^q$ is not assumed small, we can realize $\mathfrak{S}_{K_{0,Q}}$ as a quotient of a moduli scheme by a finite group, as above. In the same way, we obtain a modular description of $\mathfrak{S}_{K_{1,Q}}$:

**2.1.4.5.** For $K^q$ sufficiently small, the scheme $\mathfrak{S}_{K_{1,Q}}$ is an open and closed subscheme of the moduli scheme parametrizing quintuples

$$(A, \iota, \lambda, \beta^q, \gamma : \mathcal{O}_\mathcal{K}/\mathfrak{q}^{(1)} \hookrightarrow \mathbf{X}(A')_1),$$

where $(A, \iota, \lambda)$ are as in (2.1.2.2-4), $\beta^q$ is a level $K^q$-structure as in (2.1.3.3), and $\gamma$ is an embedding over $T$ of the constant group scheme $\mathcal{O}_\mathcal{K}/\mathfrak{q}^{(1)}$ in the $q$-divisible $\mathcal{O}_{\mathfrak{q}^{(1)}}$-module $\mathbf{X}(A')_1$, compatibly with the $\mathcal{O}_{\mathfrak{q}^{(1)}}$-structures, the whole assumed to satisfy the determinant condition (2.1.2.6). By Cartier duality, $\gamma$ defines an embedding $\gamma^* : \mu_q \to \mathbf{X}(A)_2$.

**Remark-Definition 2.1.4.6** The subgroup generated by $\gamma$ (resp. $\gamma^*$) is the one denoted $C_1$ (resp. $C_2$) above. Thus taking the subgroup generated by $\gamma$ and forgetting the generator defines a map of functors to $\mathfrak{S}_{K_{0,Q}}$. The quotient group $K_{0,Q}/K_{1,Q} = \prod_{w \in Q} k(w)^\times$ acts on the data $(\gamma_{1,q} \mid q \in Q)$, hence as a group of automorphisms of $\mathfrak{S}_{K_{1,Q}}$, with quotient $\mathfrak{S}_{K_{0,Q}}$. However, the datum (2.1.2.5) corresponds, not to $\gamma$, but rather to $\gamma^{-1}$ (a map from a subgroup of $\mathbf{X}(A')_1$ to $\mathcal{O}_\mathcal{K}/\mathfrak{q}^{(1)}$). Now $g \in K_{0,Q}/K_{1,Q}$ acts on the datum (2.1.2.5) by sending $\beta$ to $g \circ \beta$. It follows that, for $w \in Q$, $a \in k(w)^\times$ acts on $\gamma$ by sending $b \mapsto \gamma(b)$ to $b \mapsto \gamma(a^{-1}b)$. For $a \in k(w)^\times$, we thus define the *diamond operator* $< a > \in Aut(\mathfrak{S}_{K_{1,Q}}/\mathfrak{S}_{K_{0,Q}})$ to be the image of $a^{-1} \in K_{0,Q}/K_{1,Q}$. Then identifying $\gamma$ with the point $P = \gamma(1)$, the diamond operator $< a >$ takes $P$ to $a \cdot P$. The group of diamond operators is just $K_{0,Q}/K_{1,Q}$, most frequently when $Q$ consists of a single element.

**2.1.5.** Now let $\mathfrak{p}$ be a prime of $\mathcal{K}$ of residue characteristic $p$ at which $D^\#$ splits. Assume the rational prime $p$ is unramified in $\mathcal{K}$. Then $G(\mathbb{Q}_p)$ admits hyperspecial maximal compact subgroups. Suppose $K_p$ is a hyperspecial maximal compact. Then Kottwitz has shown that $\mathfrak{S}_K$ extends to a smooth proper scheme over $Spec\ \mathcal{O}_\mathfrak{p}$, representing the functor $\mathcal{A}_K$ defined as in 2.1.2, but with the additional hypothesis that the polarization $\lambda$ of (2.1.2.4) is of degree prime to the residue characteristic (cf. [HT2, K2]). In particular, the representation of $Gal(\overline{\mathbb{Q}}/\mathcal{K})$ on the $\ell$-adic cohomology is unramified at $\mathfrak{p}$, provided $\ell \neq p$.

As observed by Carayol (cf. [Ca]), the hypothesis that $p$ be unramified in $\mathcal{K}$ is in fact unnecessary, but the functor defined by 2.1.2 needs to be modified. Details can be found in [HT2,IV.2]

### 2.2. Modular definition of Hecke and $U$ operators.

**2.2.1.** We define $\Gamma_{1,q} \subset GL(n, \mathbb{Z}_q)$ as in §1.2. We consider the double coset $U_q = \Gamma_{1,q} \cdot t(\varpi)_1 \Gamma_{1,q} \subset GL(n, \mathbb{Q}_q)$. Here and in what follows we are identifying $GL(n, \mathbb{Q}_q) = U(D^\#)_q$ with $D^{\#,\times}_{\mathfrak{q}^{(2)}}$. The identification with $D^{\#,\times}_{\mathfrak{q}^{(1)}}$ identifies $U_q$ with

the double coset $\Gamma'_{1,q} \cdot (t(\varpi)_1)^{-1} \Gamma'_{1,q}$, where $\Gamma'_{1,q}$ consists of invertible matrices over $\mathbb{Z}_q$ congruent to $\begin{pmatrix} 1 & 0 \\ * & *_{n-1} \end{pmatrix}$ (mod $q$).

One verifies immediately that

$$(2.2.1.1) \qquad U_q = \coprod \Gamma_{1,q} \cdot t(\varpi)_1 \cdot \begin{pmatrix} 1 & \mathbf{b} \\ 0 & I_{n-1} \end{pmatrix},$$

where $\mathbf{b} = (b_1, \ldots, b_{n-1})$ runs through $(\mathbb{Z}_q/q\mathbb{Z}_q)^{n-1}$. Write $W = \mathbb{Q}_q^n$. Let $\Lambda = \mathbb{Z}_q^n \subset W$ be the standard lattice, and let $\lambda$ denote the row vector $(q^{-1}, 1, \ldots, 1)$, so that

$$\Gamma_{1,q} = \{g \in GL(W) \mid g(\Lambda) = \Lambda, g(\lambda) = \lambda \pmod{\Lambda}\}.$$

Let $\Lambda'$ be the lattice generated by $\Lambda$ and $\lambda$. Then the description (2.2.1.1) shows that, in the standard right action on lattices, we have

$$(2.2.1.2) \qquad (\Lambda, \lambda) \cdot U_q = \{(\Lambda', \lambda') \mid q\lambda' = \lambda \pmod{\Lambda}\}.$$

Now consider the correspondence $\mathbf{U}_q$ on $\mathfrak{S}_{K_1,Q} \times \mathfrak{S}_{K_1,Q}$, defined in terms of the data 2.1.4.5 as follows: Let $T$ be a scheme over $E(G, X)$ and let $x$ be a $T$-valued point of $\mathbb{S}_{K_1}^\circ$. Then $x$ corresponds to a quintuple $(A, \iota, \lambda, \beta^q, \gamma : \mathcal{O}_K/\mathfrak{q}^{(1)} \hookrightarrow \mathbf{X}(A')_1)$. Of course $\gamma$ is determined by the $\mathfrak{q}^{(1)}$-torsion point $P = \gamma(1)$. Let $A'$ be the abelian variety associated to $A$ and to the isogeny $\rho : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2 = \mathbf{X}(A)_2/Im(\gamma^*)$, as in (2.1.3.6), where $\gamma^* : \mu_q \to \mathbf{X}(A)_2$ is the Cartier dual to $\gamma$, as in 2.1.4.5. With respect to the Weil pairings defined by $\lambda$ and $\lambda'$, we have

$$(2.2.1.3) \qquad (P, \gamma^*(\zeta)) = \zeta \text{ for any } \zeta \in \mu_q.$$

We consider the set $\mathbf{G}^*$ of homomorphisms $(\gamma')^* : \mu_q \to \mathbf{X}(A')_2$ such that, for any $\zeta \in \mu_q$,

$$(2.2.1.4) \qquad (\gamma')^*(\zeta) = \rho(\tilde{P}), q\tilde{P} = \gamma^*(\zeta)$$

Then $\mathbf{U}_q$ associates to $x$ the set of quintuples

$$(2.2.1.5) \qquad \{(A', \iota', \lambda', (\beta')^q, \gamma')\}.$$

where $\gamma'$ runs through the set of maps $\mathcal{O}_K/\mathfrak{q}^{(1)} \hookrightarrow \mathbf{X}(A')_1$ such that $(\gamma')^* \in \mathbf{G}^*$.

LEMMA 2.2.2. *The correspondence $\mathbf{U}_q$ is induced by the Hecke operator $U_q$ by the standard recipe [K2, p. 393].*

*Proof.* This is clear from (2.2.1.2).

**2.2.3.** For use in §3, we make explicit the diagram (2.1.4.1) associated to the image of the point $(A', \iota', \lambda', (\beta')^q, \gamma')$ in $\mathfrak{S}_{K_0,Q}$. We have replaced $A$ by $A'$, and then $A'$ is replaced by an abelian variety $B'$, say, so that the diagram is

$$(2.2.3.1) \qquad \begin{array}{ccc} & B'' & \\ {}^{\phi}\swarrow & & \searrow^{\phi'} \\ A' & & B' \end{array}$$

satisfying the analogue of (2.1.4.2).

**2.3. $\lambda$-adic representations on the cohomology.** We return to the situation of §1. Let $SC$, $Q$, $\mathfrak{r}$, $K_{1,Q}$, and $V$ be as in 1.2. The primes in $Q$ are denoted $w$ rather than $\mathfrak{q}$, and the primes above them in $\mathcal{K}$ are denoted $w_1$ and $w_2$. As in §2.2, the identification $\Gamma_w \simeq GL(n, \mathbb{Q}_q)$ for $w \in Q$ is made via $w_2$, so that $K_{0,Q}$ is upper triangular parahoric. Let $\pi \in Coh(G, V)$, in the notation of §1, and assume:

$$(2.3.1) \qquad\qquad\qquad \pi^{K_{1,Q}} \neq 0$$

Then $M_\ell^{n-1}[\pi_f]$, defined as in (1.4.2), is an $n$-dimensional $\overline{\mathbb{Q}}_\ell$-vector space with a natural action of $Gal(\overline{\mathcal{K}}/\mathcal{K})$, which we henceforth denote $G_\mathcal{K}$. Let $E(\pi)_\lambda$ be as in 1.4, and write $M_\lambda[\pi_f]$ for $M_\lambda^{n-1}[\pi_f]$. We let

$$r_\rho^{\#}(\pi) : G_\mathcal{K} \rightarrow Aut(M_\lambda[\pi_f])$$

denote the natural action.

This is not quite the representation we want to associate to $\pi$. In [H1,p 101] we define a Hecke character $\alpha(\pi) = \xi \circ r_\mu$. More precisely, in [loc. cit.] we defined a Hecke character denoted $\nu(G\pi)$, which in the present notation should be equal to

$$\nu(G\pi) = \alpha(\pi)^{-1} \cdot | \bullet |_{\mathbf{A}}^{-\frac{n-1}{2}} ;$$

but the formula was incorrect; the correct formula is given below. We let $\alpha(\pi)_\lambda$ denote the corresponding $\lambda$-adic character of $G_\mathcal{K}$ – indeed, the field of coefficients of $\alpha(\pi)$ is contained in $E$, so this makes sense – and we let

$$(2.3.2) \qquad\qquad\qquad r_\rho(\pi) = r_\rho^{\#}(\pi) \otimes (\alpha_\pi)_\lambda,$$

acting on $M_\lambda[\pi_f]$. The correct formula for $\alpha(\pi)$ can be found in [H2, Errata to H1]:

$$(2.3.2.1) \qquad\qquad\qquad \alpha(\pi)(a) = \xi_\pi(N_{\mathcal{K}/\mathcal{K}_0}(a)),$$

where $\xi_\pi$ is the central character of $\pi$ and $\mathbb{G}_{m,\mathcal{K}_0}$ is regarded as the subgroup $Z_{GL(n,\mathcal{K}_0)} \cap G$ of $G$, the intersection taking place in $GL(n,\mathcal{K})$.

Next, we let $G_E$ denote $Gal(\overline{E}/E)$, and define

$$\rho(\pi) = Ind_{\mathcal{K}/E} r_\rho(\pi)$$

to be the induced $2n$-dimensional representation of $G_E$, acting on the $E(\pi)_\lambda$-module $Ind_{\mathcal{K}/E} M_\lambda[\pi_f]$.

For any prime $v$ of $\mathbb{Q}$ that splits in $\mathcal{K}_0$ and is unramified in $\mathcal{K}$, and such that $\pi_v$ is unramified we let $\phi_{v,\pi} : \mathbf{T}_v \rightarrow E(\pi)$ denote the character by which the local unramified Hecke algebra acts on $\pi_v^{K_v}$. Let $S(\pi)$ be the set of all such unramified primes, let $\mathbf{T}^S$ be the corresponding global Hecke algebra, as in §1.5, and let $\phi_\pi : \mathbf{T}^S \rightarrow E(\pi)$ be the corresponding character; $\phi_\pi$ gives the natural action of $\mathbf{T}^{S(\pi)}$ on $\pi^K$. If $w$ is a prime of $E$ dividing some prime in $S(\pi)$, we let

$$(2.3.3) \qquad\qquad \phi_\pi(P_w)(X) = 1 + \sum_{i=1}^{n} (-1)^i \phi(\pi)(T_{i,w}) X^i$$

in the notation of (1.5.2).

THEOREM 2.3.4. *Suppose $\pi$ contains a fixed vector for a hyperspecial maximal compact subgroup of $G_{q(\mathfrak{r})}$ (i.e., not only a fixed vector for $K_{q(\mathfrak{r})}$, as implied by (2.3.1)). Let $S^{bad}$ be the set of primes of $\mathcal{K}$ dividing primes in $Q$ or of residue characteristic $\ell$. Then the representation $r_\rho(\pi)$ is unramified outside $S^{bad}$. Moreover, for all but finitely many primes $w$ of $E$ dividing rational primes in $S(\pi)$, there is a prime $w_1$ of $\mathcal{K}$ dividing $w$ and such that the arithmetic Frobenius $Frob_{w_1}$ satisfies*

$$\phi_\pi(P_w)(Frob_{w_1}) = 0.$$

REMARK. The choice of $w_1$ in the above theorem is determined as in 1.5 by the choice of identification of $G'_w$ with a general linear group. The base change of $\pi$ to $\mathcal{K}$ is conjugate self-dual, and one verifies that, if $w_2$ is the other prime dividing $w$, then $\phi_\pi(P_w)(q^{n-1}Frob_{w_2}^{-1}) = 0$. Bear in mind also that the natural action on the cohomology of the Shimura variety is that of the Galois group of $\overline{\mathbb{Q}}$ over the reflex field, and that there is an implicit identification of the reflex field with $\mathcal{K}$.

An alternative way of phrasing this theorem is in terms of partial $L$-functions: there is a finite set $S$ of finite primes such that we have the equality of Euler products

$$(2.3.4.1) \qquad L^S(s, r_\rho(\pi)) = L^S(s - \frac{n-1}{2}, BC_{\mathcal{K}}(\pi)).$$

Here the right-hand side is the standard $L$-function with the unitary (Langlands) normalization; the superscript $^S$ indicates that factors at $S$ have been removed. As mentioned above, we have normalized $r_\rho$ to make (2.3.4.1) true; cf. the discussion on p. 100 ff. of [H1].

*Proof.* The proof of this theorem is mainly due to Kottwitz, and is contained in [K2]. Specifically, Kottwitz proves there that $r_\rho(\pi)$ is unramified outside $S^{bad}$, except possibly at primes of $\mathcal{K}$ ramified over $\mathbb{Q}$. Moreover, Kottwitz determines the characteristic polynomial of Frobenius in [K1] for almost all unramified primes. His original formulation includes an exponent for the (unknown) multiplicity. This exponent was removed by Taylor a few years later (cf. [H1, pp. 102-103]).

THEOREM 2.3.5. *Let $w \in Q$ of residue characteristic $q$. Let $w_1$ be the prime of $\mathcal{K}$ above $w$ and let $Z_w \subset G_E$ denote a decomposition group. Suppose $\pi_w$ is in case (b) of Proposition 1.4.3. Let $(\alpha, \beta_1, \ldots, \beta_{n-1})$ be the corresponding $n$-tuple of characters. Then $r_\rho(\pi)|_{Z_w}$ breaks up as a direct sum*

$$r_\rho(\pi)|_{Z_w} \xrightarrow{\sim} A \oplus B.$$

*Here $B$ is an unramified representation and the inertia subgroup $I_w$ of $Z_w$ acts on $A$ via the restriction to $I_w$ of the character associated to $\alpha$ via local class field theory.*

One of the main results of [HT2] is the generalization of Theorem 2.3.4 to determine the restriction of $r_\rho(\pi)$ to the decomposition group of an arbitrary prime of $\mathcal{K}$ not dividing $\ell$, up to semisimplification. In particular, Theorem 2.3.5 is a weak version, in a very special case, of what is proved in [HT2]. We give another proof in this special case, based on a different analysis of the bad reduction. The proof of Theorem 2.3.5 requires a lengthy detour through the theory of moduli of $p$-divisible groups, and is the subject of §3.

## 3. Regularity of certain moduli problems.

**3.0. Conventions.** We will now study the Shimura varieties $\mathfrak{S}_K$ at certain places of bad reduction. The techniques of the present section undoubtedly generalize, but for the time being we will make a series of restrictive hypotheses. The notation is as at the beginning of 2.1. We fix a level subgroup $K \subset G(\mathbf{A}_f)$ that satisfies (1.2.3), for $q(\mathfrak{r})$ sufficiently large. Let $q$ be a rational prime that splits completely in $\mathcal{K}$ and such that $K = K_q \times K^q$, with $K_q \subset G(\mathbb{Q}_q)$ a maximal compact subgroup. We assume the division algebra $D^{\#}$ to be split at all primes of $\mathcal{K}$ dividing $q$, and we identify $K_q$ as in 2.1.3 with

$$(3.0.1) \qquad\qquad \mathbb{Z}_q^{\times} \times \prod GL(n, \mathcal{O}_v),$$

the product being taken over divisors $v$ of $q$ in $E$.

We let $\mathfrak{q}$ be the divisor of $q$ in $E$ chosen in §2.1.3 and let $\mathfrak{q}^{(1)}$ and $\mathfrak{q}^{(2)}$ be the divisors of $\mathfrak{q}$ in $\mathcal{K}$, numbered as in (2.1.3.2). We identify $G'_{\mathfrak{q}}$ with $GL(n, \mathcal{K}_{\mathfrak{q}^{(2)}})$. Inside its compact open subgroup $GL(n, \mathcal{O}_{\mathfrak{q}^{(2)}})$, which we identify with $GL(n, \mathbb{Z}_q)$, we define the open subgroups $\Gamma_{0,q}$ and $\Gamma_{1,q}$ as in §1.2. Similarly, we define $K_{0,q}$ and $K_{1,q}$ as subgroups of $G_q$. We let

$$K_0(\mathfrak{q}) = K_{0,q} \times K^q; \quad K_1(\mathfrak{q}) = K_{1,q} \times K^q$$

be the corresponding subgroups of $K$. There is a natural isomorphism

$$(3.0.2) \qquad\qquad \mathbb{F}_q^{\times} \xrightarrow{\sim} K_0(\mathfrak{q})/K_1(\mathfrak{q}) \cong Aut(\mathfrak{S}_{K_1}(\mathfrak{q})/\mathfrak{S}_{K_0}(\mathfrak{q})).$$

We denote by $< a > \in Aut(\mathfrak{S}_{K_1}(\mathfrak{q})/\mathfrak{S}_{K_0}(\mathfrak{q}))$ the image of $a \in \mathbb{F}_q^{\times}$ under the isomorphism (3.0.2). We will later have occasion to consider subgroups intermediate between $K_0(\mathfrak{q})$ and $K_1(\mathfrak{q})$.

On the other hand, we will be studying the reduction of Shimura varieties over $\mathcal{O}_{\mathfrak{q}^{(1)}} \subset \mathcal{K}_{\mathfrak{q}^{(1)}}$. We will identify $\mathcal{O}_{\mathfrak{q}^{(1)}}$ with $\mathcal{O}_{\mathfrak{q}}$, which is isomorphic to $\mathbb{Z}_q$, and write $S_{\mathfrak{q}} = Spec(\mathcal{O}_{\mathfrak{q}})$. N.B.: the divisors in $\mathcal{K}$ of $\mathfrak{q}$ used in the geometry and in the group theory are conjugate to one another! This is an inevitable consequence of our choice of Shimura datum.

Since the prime $\mathfrak{q}$ will be fixed throughout this discussion, we will write $K_0 = K_0(\mathfrak{q})$ and $K_1 = K_1(\mathfrak{q})$ when there is no danger of confusion. We write $\mathfrak{S}_{K_1}$ and $\mathfrak{S}_{K_0}$ for the corresponding Shimura varieties over $E(G, X)$. We let $\mathcal{O}_q = \mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{Z}_q$, so that $\mathcal{O}_q \cong \prod_{v|q} \mathcal{O}_v$.

**3.1. Modular interpretations of $\mathfrak{S}_K$ and $\mathfrak{S}_{K_0}$.** As indicated in §2.1.5, the Shimura variety $\mathfrak{S}_K$ has good reduction at all primes of $\mathcal{K}$ dividing $q$. Let $\mathbb{S}_K$ denote a smooth model of $\mathfrak{S}_K$ over $S_{\mathfrak{q}}$ and let $p = p_K : \mathbb{S}_K \to S_{\mathfrak{q}}$ be the natural map. Under our hypothesis that $K$ satisfies (1.2.4) for $q(\mathfrak{r})$ sufficiently large, we have seen in 2.1.5 that $p$ is a smooth projective morphism [K2, §5]. Moreover, $\mathbb{S}_K$ is (an open and closed subscheme of) a moduli space for abelian varieties with a certain PEL type. Indeed, with our choice of $\mathfrak{q}$, the moduli problem (2.1.2.2-6) is well-defined and representable on the category of schemes over $S_{\mathfrak{q}}$, and $\mathbb{S}_K$ is the corresponding moduli scheme. Note, however, that Kottwitz' determinant condition (2.1.2.6) must be interpreted as an equality of polynomials, cf. [K2, p. 390].

Let $i : Spec(k(\mathfrak{q})) \cong Spec(\mathbb{F}_q) \hookrightarrow S_{\mathfrak{q}}$ denote the special point of $S_{\mathfrak{q}}$. The following proposition is a special case of a theorem of Wedhorn.

PROPOSITION 3.1.1 [WE1]. *Let $\bar{i} : y \to S_{\mathfrak{q}}$ be a geometric point lying over $i$, and let $\mathbb{S}_{Ky} = y \times_{S_{\mathfrak{q}}} \mathbb{S}_K$ denote the corresponding geometric special fiber. The ordinary locus is open and dense in every irreducible component of $\mathbb{S}_{Ky}$.*

Here by "ordinary locus" we mean the subscheme whose geometric points correspond to ordinary abelian varieties with structure (2.1.2.2-6).

The moduli problem defining $\mathfrak{S}_{K_0}$ is rather more intricate to describe. Our conventions will be those of Rapoport and Zink in [RZ2]. We will first define the moduli problem in their notation, then provide a brief translation into more familiar language. Let $V = R_{\mathcal{K}/\mathbb{Q}} D^{\#}$, $\mathcal{K}_q = \mathcal{K} \otimes_{\mathbb{Q}} \mathbb{Q}_q$, $E_q = E \otimes_{\mathbb{Q}} \mathbb{Q}_q$, $V_q = V \otimes_{\mathbb{Q}} \mathcal{K}_q$. Let $q^{(1)}$ and $q^{(2)}$ denote the two primes of $\mathcal{K}_0$ dividing $q$, and let

$$(3.1.2) \qquad \mathcal{K}_q = \mathcal{K} \otimes_{\mathbb{Q}} \mathbb{Q}_q \xrightarrow{\sim} \mathcal{K}_{q^{(1)}} \times \mathcal{K}_{q^{(2)}}$$

be the natural decomposition, with $\mathcal{K}_{q^{(i)}} \cong \prod_{v|q^{(i)}} \mathcal{K}_v$, $i = 1, 2$, the product taken over primes of $\mathcal{K}$. There is a corresponding decomposition $V_q \xrightarrow{\sim} V_{q^{(1)}} \times V_{q^{(2)}}$, and the hermitian form $< a, b > = Tr_{D^{\#}}(a\tilde{c}(b))$ (reduced trace) on $V$ induces a duality between $V_{q^{(1)}}$ and $V_{q^{(2)}}$:

$$(3.1.3) \qquad V_{q^{(2)}} \xrightarrow{\sim} Hom_{E_q}(V_{q^{(1)}}, E_q).$$

As in §2, we let $\mathfrak{q}^{(1)}$ and $\mathfrak{q}^{(2)}$ be the primes of $\mathcal{K}$ dividing $\mathfrak{q}$, with $\mathfrak{q}^{(1)}$ above $q^{(1)}$ and $\mathfrak{q}^{(2)}$ above $q^{(2)}$.

By a *lattice* in $V_{q^{(i)}}$, $i = 1, 2$, we mean a free $O_q^{\#}$-submodule. The subgroup $K_q$ fixes a pair of lattices $\Lambda_i \subset V_{q^{(i)}}$, $i = 1, 2$, that are placed in duality with respect to (3.1.3). The subgroup $K_{0,q}$ is the stabilizer in $K_q$ of a unique sublattice $\Lambda_1' \subset \Lambda_1$ of index $q$; we let $\Lambda_2' \supset \Lambda_2$ denote its dual in $V_{q^{(2)}}$. In general, if $\Lambda \subset V_{q^{(1)}}$ is a lattice, we write $\Lambda^{\perp}$ for its dual in $V_{q^{(2)}}$. Let $\mathcal{L}$ denote the collection of all pairs of lattices $(\Lambda, \Lambda^{\perp})$, with $\Lambda \subset V_{q^{(1)}}$, $\Lambda^{\perp} \subset V_{q^{(2)}}$, such that $\Lambda$ is fixed by $K_{0,q}$. Then $\mathcal{L}$ is an example of what Rapoport and Zink call a *multichain* of $\mathcal{O}_q$-lattices in $V_q$ [RZ2, Def. 3.10]. More precisely, the set $\Lambda \oplus \Lambda^{\perp} \subset V_q$ is the multichain, in the sense of [RZ2], but in the present case the two notions are equivalent.

For future reference, we let $\mathcal{L}_1$ denote the pair of lattices $\Lambda_2 \subset \Lambda_2'$ in $V_{q^{(2)}}$.

We will work in the category $AV$ of abelian varieties up to prime-to-$q$ isogeny. This is the category obtained from the category of abelian varieties by formally adjoining the inverses of all isogenies with kernel of order prime to $q$. It is a category whose sets of morphisms are naturally modules over the localization $\mathbb{Z}_{(q)}$ of $\mathbb{Z}$ at $q$. If $A$ is an abelian variety we let $A_{(q)}$ denote the corresponding object in $AV$. If $A, A'$ are abelian varieties, an isogeny $\rho : A_{(q)} \to A'_{(q)}$ is a morphism such that $n\rho$ is an isogeny from $A$ to $A'$ for some integer $n > 0$ prime to $q$. The $q$-part of the kernel of $\rho$ is well-defined and independent, up to canonical isomorphism, of the choice of representatives $A$ and $A'$ in the category of abelian varieties; we call it the $q$-*kernel* of $\rho$.

A quasi-isogeny $\rho : A_{(q)} \to A'_{(q)}$ is a diagram of the form

$$A_{(q)} \xleftarrow{r_1} A''_{(q)} \xrightarrow{r_2} A'_{(q)}$$

where the two arrows are isogenies. We define

$$height(\rho) = height(r_2) - height(r_1).$$

Let $AV_{\mathcal{O}_q^{\#}}$ denote the category whose objects are abelian varieties $A_{(q)}$ up to prime-to-$q$ isogeny, given with an embedding

$$\mathcal{O}_q^{\#} = \mathcal{O}^{\#} \otimes \mathbb{Z}_{(q)} \hookrightarrow End\, A_{(q)}.$$

The morphisms in $AV_{\mathcal{O}_q^{\#}}$ are those morphisms in $AV$ that respect the $\mathcal{O}_q^{\#}$-action. The $q$-divisible group $\mathbf{X}_q(A)$ of $A \in AV_{\mathcal{O}_q^{\#}}$ depends only on $A_{(q)}$. We define the $q$-divisible $\mathcal{O}_{\mathbf{q}^{(i)}}$-module $\mathbf{X}(A)_i = e_i \cdot \mathbf{X}_q(A)$, $i = 1, 2$, as in 2.1.3. Recall that $\mathbf{X}(A)_2$ is of height $n$ and dimension 1.

An $\mathcal{L}$-set of abelian varieties is a covariant functor from the category $\mathcal{L}$ (maps given by inclusions) to the category $AV_{\mathcal{O}_q^{\#}}$ satisfying several periodicity hypotheses [RZ2, Def. 6.5]. In the present situation, it corresponds to a pair $(A, A')$ of abelian varieties in the same isogeny class, determined up to prime-to-$q$ isogeny, with PEL structures (2.1.2.2-4) and a $K^q$-level structure. We are also given a pair of isogenies $\rho : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$, $\rho_1 : \mathbf{X}(A')_1 \to \mathbf{X}(A)_1$, in each case with kernels isomorphic over $\mathcal{O}_{\mathcal{K}}$ to $\mathcal{O}_{\mathbf{q}^{(i)}}/\mathbf{q}^{(i)}$-modules $C_i$, respectively, each of rank 1. It is assumed that $Ker(\rho_2)$ is the Cartier dual of $Ker(\rho_1)$ with respect to the polarization (cf. 2.1.3.6). We can express this by saying that there is a quasiisogeny $A \xleftarrow{r_2} A'' \xrightarrow{r_1} A'$ with $r_1$ and $r_2$ both isogenies of height $n$ of $\mathcal{O}^{\#}$-modules, such that $r_1$ (resp. $r_2$) induces an isomorphism on $\mathbf{X}(\bullet)_1$ (resp. on $\mathbf{X}(\bullet)_2$). In particular, $A'$ is uniquely determined by $C_2$, or equivalently by $\rho$, and by the condition that $A$ and $A'$ both admit actions by the same maximal order in $\mathcal{O}^{\#} \otimes \mathbb{Z}_q$. Thus we occasionally drop $A'$ from the notation.

Rapoport and Zink [RZ2, Def. 6.9] define a functor $\mathcal{A}_{K^q}(\mathcal{L})$, a point of which over the $S_q$-scheme $T$ is given by a quintuple as in (2.1.3.6). In the language of [RZ2], this consists of

3.1.4.1. *An $\mathcal{L}$-set $(A, A', \rho : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2)$ of objects of $AV_{\mathcal{O}_q^{\#}}(T)$, as above;*

3.1.4.2. *A $\mathbb{Q}$-homogeneous principal polarization $\lambda_A$ of $A$;*

3.1.4.3. *An $\mathcal{O}_{\mathcal{K},(q)}$-equivariant $K^q$-level structure*

$$\beta^q : H_1(A, \mathbf{A}_f^q) \simeq V \otimes \mathbf{A}_f^q \pmod{K^q}$$

*that respects the bilinear forms on both sides up to a constant in $(\mathbf{A}_f^q)^{\times}$.*

We also use the term $\mathcal{L}$-set to refer to the triple $(\mathbf{X}(A)_2, \mathbf{X}(A')_2, \rho)$. The above data have to satisfy the determinant condition (2.1.2.6), which is adequate because $q$ is assumed unramified in $\mathcal{K}$. As remarked in [RZ2, p. 279], we then have

PROPOSITION 3.1.5. *The functor $\mathcal{A}_{K^q}(\mathcal{L})$ is representable by a projective scheme over $S_{\mathfrak{q}}$, which we also denote $\mathcal{A}_{K^q}(\mathcal{L})$. The generic fiber $\mathcal{A}_{K^q}(\mathcal{L})_{\mathcal{K}_{\mathfrak{q}}}$ of $\mathcal{A}_{K^q}(\mathcal{L})$ contains $(\mathfrak{S}_{K_0})_{\mathcal{K}_{\mathfrak{q}}}$ as an open and closed subscheme.*

The second statement is as in §2.1.

In particular, we can define $\mathbb{S}_{K_0}$ to be the scheme-theoretic closure of $(\mathfrak{S}_{K_0})_{\mathcal{K}_{\mathfrak{q}}}$ in $\mathcal{A}_{K^q}(\mathcal{L})$. It is again an open and closed subscheme and the local properties of $\mathcal{A}_{K^q}(\mathcal{L})$ determine those of $\mathbb{S}_{K_0}$.

**3.2. Local models of $\mathfrak{S}_{K_0}$.** For any $A \in AV$ let $\mathbf{X}(A)$ denote its associated $q$-divisible group, and define $\mathbf{X}(A)_2$ as above. The determinant condition (2.1.2.6) translates into the condition that

3.2.1. *(a) The connected part $\mathbf{X}(A)_2^o$ of $\mathbf{X}(A)_2$ is a rank-one formal $\mathcal{O}_{\mathcal{K}, \mathbf{q}^{(2)}}$-module. In particular, $\mathbf{X}(A)_{2, \mathbf{q}^{(2)}}^o$ is a one-dimensional formal group.*

*(b) Let* $\mathfrak{q}' \neq \mathfrak{q}$ *be a prime of* $E$ *dividing* $q$, *and let* $\mathfrak{q}'^{,(i)}$ *be the divisor of* $\mathfrak{q}'$ *in* $\mathcal{K}$ *dividing* $q^{(i)}$, $i = 1, 2$. *Then the localization* $\mathbf{X}(A)_{\mathfrak{q}',(i)}$ *of* $\mathbf{X}(A)$ *at* $\mathcal{O}_{\mathcal{K},\mathfrak{q}',(i)}$ *is of multiplicative type (resp. étale) for* $i = 1$ *(resp.* $i = 2$).

Let $\mathbb{S}^s_{K_0}$ denote the special fiber of $\mathbb{S}_{K_0}$. Let $x$ be a geometric point of $\mathbb{S}^s_{K_0}$; i.e., a morphism $Spec(\overline{\kappa}) \to \mathbb{S}_{K_0}$ of $S_\mathfrak{q}$-schemes, where $\overline{\kappa}$ is the algebraic closure of the residue field $k(\mathfrak{q})$ of $\mathcal{O}_{\mathfrak{q}^{(1)}}$, which we identify with $\mathcal{O}_\mathfrak{q}$. Rapoport and Zink have shown that the infinitesimal structure of the moduli space $\mathcal{A}_{K^q}(\mathcal{L})$, and hence of the open and closed subscheme $\mathbb{S}_{K_0}$, is determined in a neighborhood of $x$ by the deformation theory of the $q$-divisible $\mathcal{O}_{\mathcal{K},(q)}$-module(s) attached to $x$.

We follow the discussion in [RZ2,6.12], to which we refer for a discussion of the crystalline data. The point $x$ corresponds to a quintuple

$$(A_x, \iota_x, \lambda_x, \beta^q_x, \rho_x : \mathbf{X}(A_x)_2 \to \mathbf{X}(A'_x)_2)$$

of data (2.1.3.6). Let $K_0$ be the field of fractions of the Witt ring $W(\overline{\kappa})$, with canonical Frobenius operator $\sigma$, and write $\overline{S}_\mathfrak{q} = Spec(W(\overline{\kappa}))$. Let $N$ be the isocrystal associated to $A_x$ by the covariant Dieudonné functor. There is an isomorphism of polarized $\mathcal{K} \otimes K_0$-modules

$$(3.2.2) \qquad N \xrightarrow{\sim} V \otimes K_0$$

The crystalline Frobenius operator $\phi$ on $N$ corresponds under this isomorphism to the operator $b \otimes \sigma$ for a (unique) $b \in G(K_0)$. To these data we can then associate a (crystalline) PEL type

$$(3.2.3) \qquad (*, \mathcal{O}_{\mathcal{K},(q)}, V, b, \mu, \mathcal{L}),$$

where $\mu$ is the minuscule cocharacter of $G$ corresponding to the Hodge decomposition (2.1.2.1). The data (3.2.3) determine a formal scheme $\check{\mathcal{M}}$ which is a moduli space, on the category $Nilp_{S_\mathfrak{q}}$ of $S_\mathfrak{q}$ schemes on which $q$ is nilpotent, for $\mathcal{L}$-sets $(\mathbf{X}, \mathbf{X}')$ of $q$-divisible $\mathcal{O}_{\mathcal{K},(q)}$-module(s) with PEL structure analogous to (2.1.2.2-4), endowed with quasi-isogenies

$$(3.2.4) \qquad \rho : (\mathbf{X}, \mathbf{X}') \longrightarrow (\mathbf{X}(A_x), \mathbf{X}(A'_x)).$$

For any such $\mathcal{L}$-set $(\mathbf{X}, \mathbf{X}')$ of $q$-divisible $\mathcal{O}_{\mathcal{K},(q)}$-modules, there is a decomposition

$$(3.2.5) \qquad \mathbf{X} \xrightarrow{\sim} \mathbf{X}_1 \times \mathbf{X}_2$$

and a quasi-isogeny

$$(3.2.6) \qquad \lambda_2 : \mathbf{X}_2 \to \widehat{\mathbf{X}_1}$$

corresponding to (3.1.2) and (3.1.3), respectively, with compatible decomposition and quasi-isogeny for the primed $q$-divisible group. The map $\lambda_2$ identifies the action of $\mathcal{O}_{\mathcal{K},q^{(2)}}$ on the left-hand side with that of $\mathcal{O}_{\mathcal{K},q^{(1)}}$ on the right, via the isomorphism (complex conjugation) $c : \mathcal{K}_{q^{(2)}} \to \mathcal{K}_{q^{(1)}}$. Moreover, these structures, together with the $\mathcal{O}_{\mathcal{K},(q)}$-action and the pair of quasi-isogenies

$$\rho_1 : \mathbf{X}_1 \to \mathbf{X}(A_x)_1; \quad \rho_2 : \mathbf{X}_2 \to \mathbf{X}(A_x)_2,$$

deduced from (3.2.4), determine the object $(\mathbf{X}, \mathbf{X}')$ uniquely. Here $\mathbf{X}(A_x)_1$ is defined as in 2.1.3 and $\mathbf{X}(A_x)_2$ is defined analogously.

We state a weak version of the Rapoport-Zink theorem on $p$-adic uniformization [RZ2, Theorem 6.23].

THEOREM 3.2.7. *Let $x$ be a geometric point of $\mathbb{S}^s_{K_0}$. There is a locally closed connected subscheme $T(x) \subset \mathbb{S}_{K_0}$ containing the image of $x$, and a connected formal subscheme $\breve{\mathcal{M}}^0 \subset \breve{\mathcal{M}}$, admitting an étale morphism*

$$\Theta : \breve{\mathcal{M}}^0 \longrightarrow \hat{T}(x)$$

*of formal schemes over $Spf(W(\overline{\kappa}))$. Here $\hat{T}(x)$ is the formal completion of $\mathbb{S}_{K_0}$ along $T(x)$.*

The statement in [RZ2] concerns the whole isogeny class containing $x$, which is a union, generally infinite, of closures of subschemes of the form $T(x)$. We are only concerned with the étale local structure of $\mathbb{S}_{K_0}$ in a neighborhood of $x$, hence the above formulation suffices for our purposes.

It follows that the nature of the singularities of of $\mathbb{S}^s_{K_0}$, if any, is encoded by the singularities of the connected components of the formal scheme $\breve{\mathcal{M}}$. Let $X^*(G)$ denote the group of $\mathbb{Q}_p$-rational characters of $G$. Rapoport and Zink construct a natural morphism $\chi : \breve{\mathcal{M}} \to \Delta$ of formal schemes over $Spf(W(\overline{\kappa}))$ where $\Delta$ is the constant group scheme over $Spf(W(\overline{\kappa}))$ [RZ2,3.52] associated to the discrete group $Hom(X^*(G), \mathbb{Z})$ (homomorphisms of abelian groups). By construction $\chi$ is a fibration. It can be defined by replacing the group $G$ by its abelianization $G^{ab}$ and by replacing the data $(b, \mu)$ in (3.2.3) by their images in $G^{ab}$, cf. the proof of Prop. 1.21 in [RZ2]; however, we will not use this fact.

Now we return to the decomposition (3.2.5). The map

(3.2.8)                          $(\mathbf{X}, \mathbf{X}'; \rho) \mapsto (\mathbf{X}_2, \mathbf{X}'_2; \rho_2)$

defines a morphism

$$f_1 : \breve{\mathcal{M}} \longrightarrow \breve{\mathcal{M}}_1$$

where $\breve{\mathcal{M}}_1$ is the moduli space of EL type, in the terminology of [RZ2,§3] (see below), parametrizing $\mathcal{L}_1$-sets $(\mathbf{X}_1, \mathbf{X}'_1)$ of $q$-divisible $\mathcal{O}_{\mathcal{K}, q^{(1)}}$-modules of height n and dimension 1, rigidified by a quasi-isogeny $\rho_1$ to the fixed $\mathcal{L}_1$-set $(\mathbf{X}_2(A_x), \mathbf{X}_2(A'_x))$. Here by $\mathcal{L}_1$-set we just mean a pair $(\mathbf{X}_2, \mathbf{X}'_2)$ as above connected by an isogeny $\mathbf{X}_2 \to \mathbf{X}'_2$ of degree $q$ (cf. the definitions following (3.1.3)). By the discussion preceding Theorem 3.2.7, we see that this morphism is just the forgetful map "forget $\lambda_2$". The set of possible $\lambda_2$'s is a principal homogeneous space under $\mathbb{Q}^\times_q / \mathbb{Z}^\times_q$, thus the map $f_1$ is étale. Moreover, if $\breve{\mathcal{M}}^0 \subset \breve{\mathcal{M}}$ is a connected component, $f$ induces an isomorphism $\breve{\mathcal{M}}^0 \xrightarrow{\sim} f(\breve{\mathcal{M}}^0)$. Thus we have

COROLLARY 3.2.9. *In Theorem 3.2.7 $\breve{\mathcal{M}}$ can be replaced by $\breve{\mathcal{M}}_1$.*

Now we are reduced to studying the singularities of $\breve{\mathcal{M}}_1$. We begin by recalling the definition of the EL data defining $\breve{\mathcal{M}}_1$. We are given the $\mathbb{Q}_p$-algebra $\mathcal{K}_{q^{(2)}}$, its maximal order $\mathcal{O}_{\mathcal{K}, q^{(2)}}$, the free rank $n$ $\mathcal{K}_{q^{(2)}}$-module $V_{q^{(2)}}$, and its rank one submodule $V_{0, q^{(2)}}$. To these data is associated the group $G_2 = GL(n)/\mathcal{K}_{q^{(2)}}$. We are also given the crystalline Frobenius $\phi_2 = \phi|V_{q^{(2)}} \otimes K_0$ with respect to the isomorphism (3.2.5), and hence the image $b_2$ of the element $b \in G(K_0)$ under the natural map $G \to G_2$ (forget the similitude factor).

However, the structure of possible singularities is independent of $b_2$. More precisely, Rapoport and Zink define a *local model* $\hat{\mathbf{M}}^{loc}$, a projective formal scheme over $Spf(W(\overline{\kappa}))$ depending only on the data $(\mathcal{K}_{q^{(1)}}, \mathcal{O}_{\mathcal{K},q^{(2)}}, V_{q^{(2)}}, V_{1,q^{(2)}})$, for which the following proposition holds:

**PROPOSITION 3.2.10.** *([RZ2,Prop. 3.33]) Let $x \in \breve{\mathcal{M}}_1$ be a closed point. Then $x$ has a pointed étale neighborhood $(\mathcal{U}, y) \to (\breve{\mathcal{M}}_1, x)$ which is formally étale over $\hat{\mathbf{M}}^{loc}$.*

The construction underlying this proposition will be outlined in the course of the proof of Lemma 3.3.2. We copy the definition of the functor $\mathbf{M}^{loc}$ from [RZ2, Definition 3.27]. Recall that we identify $\mathcal{O}_{\mathcal{K},q^{(1)}}$ with $\mathcal{O}_q$.

**DEFINITION 3.2.11.** *A point of $\mathbf{M}^{loc}$ with values in an $\mathcal{O}_q$-scheme $T$ is given by*
*(a) an $\mathcal{L}_1$-set of locally free rank one $\mathcal{O}_T$-modules $t_{\Lambda_2} \to t_{\Lambda_2'}$ and*
*(b) a morphism of $\mathcal{L}_1$-sets*

$$\phi' : \Lambda_2' \otimes_{\mathbb{Z}_q} \mathcal{O}_T \to t_{\Lambda_2'}; \quad \phi : \Lambda_2 \otimes_{\mathbb{Z}_q} \mathcal{O}_T \to t_{\Lambda_2},$$

*Both $\phi'$ and $\phi$ are surjective, and the obvious diagram is assumed commutative.* We let $\hat{\mathbf{M}}^{loc}$ denote the $q$-adic completion of $\mathbf{M}^{loc} \times_{S_q} Spec(W(\overline{\kappa}))$.

**PROPOSITION 3.2.12.** *Let $s$ be a closed point of $\hat{\mathbf{M}}^{loc}$, and let $\hat{\mathbf{M}}_s^{loc}$ denote the formal completion of $\hat{\mathbf{M}}^{loc}$ along $s$. Suppose $\hat{\mathbf{M}}^{loc}$ is singular at $s$. Then there are a complete local smooth $W(\overline{\kappa})$-algebra $S$ and an isomorphism*

$$\hat{\mathbf{M}}_s^{loc} \xrightarrow{\sim} Spf(S[[u,v]]/(uv-q))$$

*of formal schemes over $Spf(W(\overline{\kappa}))$.*

*In particular, $\hat{\mathbf{M}}^{loc}$ is regular and flat over $Spf(W(\overline{\kappa}))$, and its special fiber has only singularities of the form (smooth) × (ordinary double point).*

*Proof.* The proof proceeds by making the local equations for $\mathbf{M}^{loc}$ explicit. Our approach is similar to that of de Jong in [dJ]. We work over $W(\overline{\kappa})$. If $T$ is a $W(\overline{\kappa})$-scheme, the points of $\mathbf{M}^{loc}$ with values in $T$ are given by the data (a) and (b) of Definition 3.2.11. In particular, if $T$ is in characteristic zero then the homomorphism $t_{\Lambda_2} \to t_{\Lambda_2'}$ is necessarily an isomorphism.

We choose bases $\{e, e_1, \ldots, e_{n-1}\}$ and $\{f, f_1, \ldots, f_{n-1}\}$ of $\Lambda_2'$ and $\Lambda_2$, respectively, so that

$$e = qf, \ e_i = f_i, \quad i = 1, \ldots, n-1.$$

These bases identify the projectivizations $\mathbb{P}(\Lambda_2')$ and $\mathbb{P}(\Lambda_2)$ with projective space $\mathbb{P}_{K_0}^{n-1}$. Let $\{X, X_1, \ldots, X_{n-1}\}$ and $\{Y, Y_1, \ldots, Y_{n-1}\}$ be the corresponding sets of homogeneous coordinates. The quotients $t_{\Lambda_2'}$ of $\Lambda_2' \otimes_{\mathbb{Z}_q} \mathcal{O}_T$ defines a $T$-valued point of $\mathbb{P}(\Lambda_2')$; likewise $t_{\Lambda_2}$ defines a $T$-valued point of $\mathbb{P}(\Lambda_2)$. The compatibility condition is that these two points are identified to the same point in $\mathbb{P}_{K_0}^{n-1}$. Assume $T = Spec(R)$ to be affine and let $(x, x_1, \ldots, x_{n-1})$ and $(y, y_1, \ldots, y_{n-1})$ be their expressions in the homogeneous coordinates defined above, with the $x$'s and $y$'s in $R$; then the condition is that $(qx, x_1, \ldots, x_{n-1})$ and $(y, y_1, \ldots, y_{n-1})$ are proportional. In other words, the matrix

(3.2.12.1)
$$\begin{pmatrix} qx & x_1 & \ldots & x_{n-1} \\ y & y_1 & \ldots & y_{n-1} \end{pmatrix}$$

is of rank one.

The 2-by-2 minors of (3.2.12.1) provide a set of homogeneous equations for the scheme $\mathbf{M}^{loc}$. The assertions of the proposition follow easily from the form of these equations, as we will see in a moment. We note for future reference the following fact:

REMARK 3.2.12.2. *The open subscheme* $\mathbf{M}^{loc,\acute{e}} \subset \mathbf{M}^{loc}$, *defined by the condition that the homomorphism* $t_{\Lambda_2} \to t_{\Lambda_2'}$ *be an isomorphism, is non-singular.*

It suffices to consider a geometric point $z$ of $\mathbf{M}^{loc,\acute{e}}$ in characteristic $q$. Let $R$ be a complete local ring with residue field $\overline{\kappa}$ and let the $x$'s and $y$'s be local homogeneous coordinates near $z$, as above; we may assume the $x$'s and $y$'s lie in $R$. The condition $z \in \mathbf{M}^{loc,\acute{e}}$ corresponds to the condition that the rows of (3.2.12.1) be proportional with a factor of proportionality $\alpha$ invertible in $R$. Consider the inhomogeneous coordinates. There are essentially three possibilities:

(a) $x_1$ is a unit, $y_1$ is a unit. Let

$$x_j' = \frac{x_j}{x_1}, \ y_j' = \frac{y_j}{y_1}, j > 1, \ y' = \frac{y}{y_1}, \ x' = \frac{x}{x_1}.$$

The inhomogeneous equations are

$$y' = x'; \ x_j' = y_j', j > 1,$$

and these are always smooth.

(b) $y_i$ is a unit for some $i \geq 1$. Without loss of generality we may assume $i = 1$. Then the invertibility of $\alpha$ implies $x_1$ is also a unit, and we are reduced to case (a).

(c) $y$ is a unit. This is impossible because $y = \alpha q x$ with $\alpha$ invertible and $x \in R$.

It is case (c) that gives rise to the singularities of $\mathbf{M}^{loc}$ outside $\mathbf{M}^{loc,\acute{e}}$.

The following theorem appears to be the natural generalization to higher dimensions of the well-known theorem of Deligne and Rapoport on the local structure of the modular curve $X_0(q)$.

THEOREM 3.2.13. *The moduli scheme* $\mathbb{S}_{K_0}$ *is regular and flat over* $S_\mathfrak{q}$, *and its special fiber is a union of smooth divisors with normal crossings. Moreover, if $x$ is a singular point of the special fiber then the formal completion* $\mathbb{S}_{K_0,x}$ *is isomorphic to a formal scheme of the form* $Spf(S[[u,v]]/(uv-q))$, *where $S$ is a complete local smooth* $\mathcal{O}_\mathfrak{q}$-*algebra.*

*Proof.* The properties in question are local in the étale topology. Thus it suffices to verify the theorem after base change to $W(\overline{\kappa})$. The Theorem is thus a consequence of Theorem 3.2.7, Corollary 3.2.9, Proposition 3.2.10, and Proposition 3.2.12.

The following proposition is a strengthening of Proposition 3.1.1.:

PROPOSITION 3.2.14. *Let* $\overline{i} : y \to S_\mathfrak{q}$ *be a geometric point lying over $i$, and let* $\mathbb{S}_{K_0,y} = y \times_{S_\mathfrak{q}} \mathbb{S}_{K_0}$ *denote the corresponding geometric special fiber. The ordinary locus is open and dense in every irreducible component of* $\mathbb{S}_{K_0,y}$.

*Proof.* First note that the forgetful morphism $\mathbb{S}_{K_0,y} \to \mathbb{S}_{K,y}$ is finite, i.e. has finite fibers. Indeed, let $x$ be a geometric point of $\mathbb{S}_{K,y}$, corresponding to an abelian variety $A_x$; then the points of $\mathbb{S}_{K_0,y}$ above $x$ correspond to the rank 1 $\mathcal{O}_{\mathfrak{q}^{(2)}}/\mathfrak{q}^{(2)}$-subgroup schemes $C_2 \subset A_x[\mathfrak{q}^{(2)}]$. But the connected part of the finite flat group scheme $A_x[\mathfrak{q}^{(1)}]$ is of dimension one. It follows easily that the set of possible $C_2$'s is finite.

On the other hand, it follows from Theorem 3.2.13 that every irreducible component $\mathbb{S}_{K_0,y}$ is of dimension $n-1$. Combining these two facts with Proposition 3.1.1,

we see that no irreducible component of $\mathbb{S}_{K_0,y}$ lies entirely over the complement of the ordinary locus in $\mathbb{S}_{K,y}$. The assertion now follows from Proposition 3.1.1.

REMARKS. The results of this section were obtained in 1996. Analogous results for Iwahori level were known to Rapoport previously (cf. [R] which, however, only treats places where the division algebra is ramified). General parahoric levels and general signatures have since been treated by Görtz [G], and Theorem 3.2.13 is now regarded as the trivial case.

Counterexamples due to Stamm show that the analogue of Proposition 3.2.14 is not valid in the generality of [We1]. However, the present case has also been considered in [We2], where Wedhorn obtains a congruence formula related to the one proved in section 4.2, below.

## 3.3. Local models of $\mathfrak{S}_{K_1}$.

**3.3.1.** Let $(A^{univ}, A'^{,univ}, \rho^{univ} : \mathbf{X}(A^{univ})_2 \to \mathbf{X}(A'^{,univ})_2)$ denote the universal $\mathcal{L}$-set (3.1.4.1) of abelian varieties over the moduli scheme $\mathcal{A}_{K^q}(\mathcal{L})$. Then the kernel of $\rho^{univ}$ is a finite flat group scheme $C_2^{univ}$ of order $q$, isomorphic over $\mathcal{O}_{\mathcal{K}}$ to an $\mathcal{O}_{\mathfrak{q}^{(2)}}/\mathfrak{q}^{(2)}$-module of rank 1 (3.1.4). We use the same notation to denote the pullback of $C_2^{univ}$ to $\mathbb{S}_{K_0}$ via the natural inclusion. Similarly, we let $C_1^{univ}$ denote the kernel of the homomorphism $\rho'^{,univ} : \mathbf{X}(A'^{,univ})_1 \to \mathbf{X}(A^{univ})_1$ described in 2.1.3.6, so that $C_1^{univ}$ and $C_2^{univ}$ are canonically Cartier dual (2.1.4.4).

The property "$C_2$ is étale" is relatively representable over the functor $\mathcal{A}_{K^q}(\mathcal{L})$ and defines an open subscheme $\mathcal{A}_{K^q}(\mathcal{L})_{\acute{e}}^o$ [cf. [DR,§V.1]. Likewise, the property "$C_1$ is étale", equivalent to the property "$C_2$ is locally isomorphic to $\mu_q$ in the étale topology", defines an open subscheme $\mathcal{A}_{K^q}(\mathcal{L})_m^o$ of $\mathcal{A}_{K^q}(\mathcal{L})$. It follows from Proposition 3.2.14 that

**3.3.1.1.** *The union* $\mathcal{A}_{K^q}(\mathcal{L})^o = \mathcal{A}_{K^q}(\mathcal{L})_{\acute{e}} \cup \mathcal{A}_{K^q}(\mathcal{L})_m^o$ *is open and dense in* $\mathcal{A}_{K^q}(\mathcal{L})$.

We let $\mathbb{S}_{K_0}^o$, $\mathbb{S}_{K_0,\acute{e}}^o$, and $\mathbb{S}_{K_0,m}^o$ denote the corresponding subschemes of $\mathbb{S}_{K_0}$. Note that $\mathbb{S}_{K_0}^o$ contains the ordinary locus but is in general larger.

Let $\overline{\mathbb{S}}_{K_0}$ denote the special fiber of $\mathbb{S}_{K_0}$, $X_m^o = \overline{\mathbb{S}}_{K_0} \cap \mathbb{S}_{K_0,m}^o$, $X_{\acute{e}}^o = \overline{\mathbb{S}}_{K_0} \cap \mathbb{S}_{K_0,\acute{e}}^o$, and let $X_m$ and $X_{\acute{e}}$ denote the closures in $\overline{\mathbb{S}}_{K_0}$ of $X_m^o$ and $X_{\acute{e}}^o$, respectively. It follows from (3.3.1.1) that

$$(3.3.1.2) \qquad\qquad \overline{\mathbb{S}}_{K_0} = X_m \cup X_{\acute{e}}.$$

We will need the following strengthening of Theorem 3.2.13:

LEMMA 3.3.2. *Let $x$ be a singular point of the special fiber of $\mathbb{S}_{K_0}$, and let*

$$(A_x, A_x', \rho_x : \mathbf{X}(A_x)_2 \to \mathbf{X}(A_x')_2)$$

*be the corresponding $\mathcal{L}$-set of objects of $AV_{\mathcal{O}_{\mathfrak{q}}^{\#}}$. Then $Ker\, \rho_x$ is isomorphic to the unipotent group scheme $\alpha_q$.*

*Proof.* First, suppose $Ker\, \rho_x$ is of multiplicative type. Since $A_x[\mathfrak{q}^{(2)}]$ is one-dimensional, it follows that $x$ is in the ordinary locus and $Ker\, \rho_x$ is the full $q$-torsion subgroup $\mathbf{X}(A)_2^o[q]$ of the formal group $\mathbf{X}(A)_2^o$ (cf. 3.2.4). Let $(\overline{\mathbb{S}}_K)^o$, resp. $(\overline{\mathbb{S}}_{K_0})^o$ denote the ordinary locus of $\overline{\mathbb{S}}_K$, resp. $\overline{\mathbb{S}}_{K_0}$. The former is an open and closed

subscheme of the moduli space $\mathcal{A}_K$ for triples $(A, \lambda, \beta)$ as in 3.1.4, and the functor "forget $\rho : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$" from $\mathcal{A}_{K^q}(\mathcal{L})$ to $\mathcal{A}_K$ restricts to the canonical map $f : X_m^o \to (\overline{\mathbb{S}}_K)^o$. Then the map $f$ admits a section:

$$(A, \lambda, \beta) \mapsto (A, A', \rho, \lambda, \beta),$$

where $\rho : \mathbf{X}_1 \to \mathbf{X}_2 / \mathbf{X}(A)_2^o[q]$ is the tautological map and $A'$ is determined by $A$ and $Ker(\rho)$ as in the discussion preceding 3.1.4.1. Since $\overline{\mathbb{S}}_K$ is smooth, it follows that $X_m^o$ is smooth, hence $x$ is a smooth point.

To treat the étale case we need to recall briefly how Proposition 3.2.10 is proved. Let $(\mathbb{X}_2, \mathbb{X}_2'; \rho_2)$ be the universal rigidified $\mathcal{L}_1$-set of $q$-divisible $\mathcal{O}_{K,q^{(2)}}$-modules over $\breve{\mathcal{M}}_1$, in the notation of (3.2.8). Let $Nilp_{W(\overline{\kappa})}$ be the category of $W(\overline{\kappa})$ schemes on which $q$ is locally nilpotent. Let $\mathcal{N}$ be the functor on $Nilp_{W(\overline{\kappa})}$ such that, for $S \in Nilp_{W(\overline{\kappa})}$, $\mathcal{N}(S)$ is given by triples $((\mathbf{X}_2, \mathbf{X}_2'); \rho_2; (\gamma_2, \gamma_2'))$. Here $((\mathbf{X}_2, \mathbf{X}_2'); \rho_2) \in \breve{\mathcal{M}}_1(S)$ and

$$(\gamma_2, \gamma_2') : (M(\mathbf{X}_2), M(\mathbf{X}_2')) \xrightarrow{\sim} (\Lambda_2, \Lambda_2') \otimes_{\mathbb{Z}_q} \mathcal{O}_S$$

is an isomorphism of $\mathcal{L}_1$-sets of $\mathcal{O}_S$-modules, where $M(\bullet)$ denotes the crystal associated to the $q$-divisible $\mathcal{O}_{K,q^{(2)}}$-module $\bullet$. Then $\mathcal{N}$ is representable by a formal scheme. The forgetful functor defines a formally smooth morphism $\mathcal{N} \to \breve{\mathcal{M}}_1$.

Consider the morphism $\psi$ of functors that to the triple $((\mathbf{X}_2, \mathbf{X}_2'); \rho_2; (\gamma_2, \gamma_2'))$ associates the composite morphism of $\mathcal{L}_1$-sets:

(3.3.2.1)        $(\Lambda_2, \Lambda_2') \otimes_{\mathbb{Z}_q} \mathcal{O}_S \xrightarrow{\sim} (M(\mathbf{X}_2), M(\mathbf{X}_2')) \to (Lie(\mathbf{X}_2), Lie(\mathbf{X}_2')).$

Here the first arrow is given by $(\gamma_2, \gamma_2')^{-1}$, while the second morphism is given by the theory of Grothendieck and Messing [Me]. Then $\psi$ defines a morphism $\psi : \mathcal{N} \to \hat{\mathbf{M}}^{loc}$ (cf. [RZ2], p. 90).

Let $x \in \breve{\mathcal{M}}_1$ be a closed geometric point, and assume the corresponding morphism $\mathbf{X}_{2,x} \to \mathbf{X}_{2,x}'$ is étale. We need to show that $x$ is a smooth point of $\breve{\mathcal{M}}_1$. Rapoport and Zink consider a certain class of sections $s$ of $\mathcal{N} \to \breve{\mathcal{M}}_1$ on pointed étale neighborhoods $(\mathcal{U}, y)$ of $(\breve{\mathcal{M}}_1, x)$, which are called *rigid of the first order* [RZ2, p. 91]. They show that if $s : (\mathcal{U}, y) \to \breve{\mathcal{M}}_1$ is rigid of the first order, then $\psi \circ s : \mathcal{U} \to \hat{\mathbf{M}}^{loc}$ is formally étale in a Zariski open neighborhood of $y$ ([RZ2, Prop. 3.33]). On the other hand, it is immediate from the definitions that if $\mathbf{X}_{2,x} \to \mathbf{X}_{2,x}'$ is étale, then $s(y) \in \hat{\mathbf{M}}^{loc,é}$, defined as the $q$-adic completion of $\mathbf{M}^{loc,é} \times_{Spec(\mathcal{O}_q)} W(\overline{\kappa})$. But Remark 3.2.12.2 implies that $\hat{\mathbf{M}}^{loc,é}$ is smooth. This implies that $x$ is a smooth point, and completes the proof of Lemma 3.3.2.

PROPOSITION 3.3.3. *The subschemes $X_m$ and $X_é$ are smooth divisors on $\mathbb{S}_{K_0}$.*

*Proof.* Let $x$ be a singular point of $\overline{\mathbb{S}}_{K_0}$ contained in $X_m$. We replace $\overline{\mathbb{S}}_{K_0}$ by an appropriate étale neighborhood of $x$, which we may take to be isomorphic to $Spec(R)$, $R = S[u, v]/(uv - q)$, with notation as in Theorem 3.2.13. It follows from Theorem 3.2.13 that $x$ lies on the intersection of two smooth irreducible components of the inverse image in $Spec(R)$ of $X_m$, say $D_1$ and $D_2$. Following a suggestion of de Jong, we use the Tate-Oort theory [TO] to describe the group scheme of prime order $C_1^{univ}$ in a neighborhood of $x$. We may assume $u$ to be a local parameter for $D_1$ and $v$ a local parameter for $D_2$. Let $C$ denote the pullback of $C_1^{univ}$ to $Spec(R)$. Since $R$

is affine it follows from the main theorem of [TO] that $C$ is determined by a pair of elements $x, y \in R$ with $xy = q$, and that $C$ is étale on the subset of $Spec(R)$ where $x$ is invertible. But our hypothesis is that $C_2^{univ}$ is multiplicative, hence $C_1^{univ}$ is étale, on the non-singular locus of $D_1 \cup D_2$. Since $\mathbb{S}_{K_0}$ is regular, the subvariety defined by the vanishing of $x$ is purely of codimension one. Thus $x$ is invertible in $R$, and $C$ is étale on $Spec(R)$. But then it follows from Lemma 3.3.2 that $x$ is not a singular point. The argument for $X_\acute{e}$ is analogous.

REMARK 3.3.4. Let $R$ be a $\mathbb{Z}_q$ algebra. The Tate-Oort theory actually classifies group schemes of order $q$ over $Spec(R)$ by elements $a$ and $b$ in $R$ such that that $ab = w_q$, where $w_q = \nu q$ for a certain explicit unit $\nu$ in $\mathbb{Z}_q$. Our parameters above are $x = \nu^{-1}a$, $y = b$. The distinction is important for the group structure but not for the local geometry. Thus we let $G_R(x, y)$ denote the group over $Spec(R)$ associated to $(a = \nu x, b = y)$ in the Tate-Oort classification. Then

$\qquad G_R(x, y) \cong G_R(x', y')$ *if and only if* $x/x'$ *is the* $(q-1)$st *power of a unit in* $R$.
The proof yields the following Corollary:

COROLLARY 3.3.5. *Let* $s$ *be a singular point of* $\overline{\mathbb{S}}_{K_0}$ *and let* $Spf(S[[u,v]]/(uv-q))$ *be a local model for the formal completion of* $\mathbb{S}_{K_0}$ *along* $s$, *in the notation of Theorem 3.2.13. Then* $s$ *lies in the intersection* $X_m \cap X_\acute{e}$, *and, up to interchanging* $u$ *and* $v$, $u$ *is a local equation for* $X_\acute{e}$ *and* $v$ *is a local equation for* $X_m$. *Moreover, in the Tate-Oort classification for the pullback* $C$ *of* $C_1^{univ}$ *to* $Spf(S[[u,v]]/(uv-q))$, *we have* $C \simeq G_R(x, y)$, *with* $R = S[[u,v]]/(uv-q)$, $x = \alpha u$, $y = \alpha^{-1}v$, *for some* $\alpha \in (S[[u,v]]/(uv-q))^\times$.

*Proof.* Only the last statement needs justification. But in the Tate-Oort theory, $x$ (resp. $y$) vanishes precisely where $C_1^{univ}$ is not étale (resp. multiplicative). Thus the variety defined by $x$ is the subset $X_\acute{e}$ where $C_2^{univ}$ is not multiplicative, so $x = au$ for some $a \in (S[[u,v]]/(uv-q))^\times$, and we conclude by recalling that $xy = q = uv$.

We now turn to the moduli scheme $\mathfrak{S}_{K_1}$. Let $C^+$ denote the complement of the zero section of $C_1^{univ}$ over the generic fiber $(\mathfrak{S}_{K_0})_{\mathcal{K}_\mathfrak{q}}$. Then $C^+$ tautologically represents the functor

$$(3.3.5.1) \qquad T \mapsto \left\{ \begin{array}{l} T\text{-valued point } f : T \to (\mathfrak{S}_{K_0})_{\mathcal{K}_\mathfrak{q}} \\ \text{plus a generator of } f^*(C_1^{univ}) \text{ over } T \end{array} \right\}$$

On the other hand, just as in the elliptic modular case (3.3.5.1) is represented by $\mathfrak{S}_{K_1}$ over $\mathcal{K}_\mathfrak{q}$ (cf. 2.1.4.5). Thus in order to obtain a regular model for $\mathfrak{S}_{K_1}$ over $S_\mathfrak{q}$ it will be enough to find a regular model for $C^+$. We do this explicitly using the Tate-Oort equations.

PROPOSITION 3.3.6. *(a) The moduli scheme* $\mathfrak{S}_{K_1}$ *has a flat model* $\mathbb{S}_{K_1}$ *over* $S_\mathfrak{q}$ *which is a regular scheme, and whose special fiber is a divisor with normal crossings, each of whose components has multiplicity prime to* $q$. *More precisely, there is a finite flat morphism* $p_1 : \mathbb{S}_{K_1} \to \mathbb{S}_{K_0}$.

*(b) The inverse image* $Y_m = p_1^{-1}(X_m)$ *is a reduced divisor on* $\mathbb{S}_{K_1}$, *while* $Y_\acute{e} = p_1^{-1}(X_\acute{e})$ *is a divisor with multiplicity* $q - 1$, *and the map* $p_1 : Y_{\acute{e},red} \to X_\acute{e}$ *is an isomorphism.*

*(c) Let* $Y_{sing}$ *denote the singular locus of* $\mathbb{S}_{K_1}$. *Then* $Y_{sing,red}$ *is smooth and all its components are of codimension 1 in the special fiber.*

*Proof.* We define $\mathbb{S}_{K_1}$ to be the normalization of $\mathbb{S}_{K_0}$ in $C^+$. All assertions are local on $\mathbb{S}_{K_0}$, so we can work over a neighborhood of a geometric point $s$ of the special fiber $\overline{\mathbb{S}}_{K_0}$. There are three cases to consider.

*Case 1:* $s \in X_m^o$. Then $C^+$ is étale over $\mathbb{S}_{K_0}$ in a neighborhood of $s$, and the claims follow immediately from Theorem 3.2.13.

*Case 2:* $s \in X_{\acute{e}}^o$. Then $C^+$ is of multiplicative type, hence is isomorphic to $\mu_q$ on an étale neighborhood of $s$. The statement about multiplicities is then obvious, as is the assertion that $p_1 : Y_{\acute{e},red} \to X_{\acute{e}}$ is an isomorphism.

*Case 3:* $s \in X_{\acute{e}} \cap X_m$. We use the notation of Corollary 3.3.5. Let $\alpha$ be the unit introduced in the statement of Corollary 3.3.5. By the remark preceding that corollary, $C \simeq G_R(u,v)$ if and only if the unit $\alpha$ is a $(q-1)$st power in $S[[u,v]]/(uv-q)$. Now the extension $(S[[u,v]]/(uv-q))[(\alpha)^{\frac{1}{q-1}}]$ is finite and étale over $S[[u,v]]/(uv-q)$. Since our theorem is local in the étale topology, we may assume $C \simeq G_R(u,v)$.

Now we recall the Tate-Oort local equations for $G_R(u,v)$. On p. 13 of [TO] are introduced parameters $U$ and $y$ such that, setting $X_1 = w_q U^{1-q}$, $X_2 = U^{q-1}$, and $Y = U^{-1}y$, and working on the formal completion of $Spec\ S[u,v]/(uv-q)$, we have

$$(3.3.7)\quad G_R(u,v) = Spf(S[Y][[X_1,X_2]]/(X_1 X_2 - w_q, Y^q - X_1 Y); \quad X_1 = \nu u, \ X_2 = v.$$

Now the factorization $Y^q - X_1 Y = Y(Y^{q-1} - X_1)$ defines an injective map from the ring on the right-hand side of (3.3.7) to the direct product of two rings $R_0$ and $R^*$, where

$$R_0 = S[[X_1, X_2]]/(X_1 X_2 - w_q); \quad R^* = S[[X_2, Y]]/(Y^{q-1} X_2 - w_q).$$

Dually, we have a surjective morphism

$$Spf(R_0) \coprod Spf(R^*) \ \to \ G_R(u,v).$$

The image of $Spf(R_0)$ corresponds to the zero section, while the image of $Spf(R^*)$ is the scheme-theoretic closure of $C^+$. It follows that $Spf(R^*)$ is a local model for $C^+$ over a neighborhood of the singular point $s$. Obviously $Spf(R^*)$ has the properties required, and the construction shows that $p_1 : Y_{e,red} \to X_{\acute{e}}$ is an isomorphism near $s$. Moreover, $Spf(R^*)$ is normal, and it follows just as in [DR,§V.2] that $Spf(R^*)$ is a formal local model for $\mathbb{S}_{K_1}$ near $s$.

This completes the proof.

The analogy with [DR] becomes clearer if we extend scalars to $\mathcal{O}_{\mathfrak{q}}[\zeta_q]$, where $\zeta_q$ denotes a primitive $q$-th root of 1. Let $\tilde{S}_{\mathfrak{q}} = Spec(\mathcal{O}_{\mathfrak{q}}[\zeta_q])$.

PROPOSITION 3.3.8. *(a) The moduli scheme $\mathfrak{S}_{K_1} \times_{\mathcal{K}_{\mathfrak{q}}} \mathcal{K}_{\mathfrak{q}}(\zeta_q)$ has a flat model $\tilde{\mathbb{S}}_{K_1}$ over $\tilde{S}_{\mathfrak{q}}$ which is a regular scheme, and whose special fiber is a reduced divisor with normal crossings. More precisely, there is a finite flat morphism $\tilde{p}_1 : \tilde{\mathbb{S}}_{K_1} \to \mathbb{S}_{K_0} \times_{S_{\mathfrak{q}}} \tilde{S}_{\mathfrak{q}}$.*

*(b) The inverse images $\tilde{Y}_{\acute{e}} = \tilde{p}_1^{-1}(X_{\acute{e}})$ and $\tilde{Y}_m = \tilde{p}_1^{-1}(X_m)$ are reduced divisors on $\tilde{\mathbb{S}}_{K_1}$. Moreover, $\tilde{Y}_m \simeq Y_m$ and $\tilde{Y}_e^o = \tilde{p}_1^{-1}(X_{\acute{e}}^o)$ is étale over $(Y_{\acute{e}}^o)_{red} \simeq X_{\acute{e}}^o$ with Galois group isomorphic to the group of diamond operators (cf. Remark-Definition 2.1.4.6).*

*(c) The model $\tilde{\mathbb{S}}_{K_1}$ can be defined as the normalization of $\mathbb{S}_{K_0} \times_{S_{\mathfrak{q}}} \tilde{S}_{\mathfrak{q}}$ in the scheme*

$$Isom_{\mathbb{S}_{K_0} \times_{S_{\mathfrak{q}}} Spec(\mathcal{K}_{\mathfrak{q}}(\zeta_q))}(\mathbb{Z}/q\mathbb{Z}, C_1).$$

*(d) Over the complement of $\tilde{Y}_m$ (resp. $\tilde{Y}_{\acute{e}}$) $\tilde{\mathbb{S}}_{K_1}$ represents the functor $Isom(\mu_q, C_1)$ (resp. $Isom(\mathbb{Z}/q\mathbb{Z}, C_1)$), and these two are canonically isomorphic over the generic fiber.*

*Proof.* The first part of the proof of Proposition 3.3.6 showed that we can take the universal group scheme $C$ over $\mathbb{S}_{K_0}$ to be étale locally isomorphic to $G_R(u,v)$. Let $R_0 = \mathbb{Z}_q[u,v]/(uv - q)$. Then $G_R(u,v)$ is the base change from $R_0$ to $R$ of $G_{R_0}(u,v)$, and it suffices to work over the base $R_0$, a local curve over $Spec(\mathbb{Z}_q) \cong S_{\mathfrak{q}}$. Then the proof of [DR,Lemma V.2.8] applies word for word to yield the result. Parts (c) and (d) are the analogues of [DR, Theorem 2.7] and [DR, Prop. 2.3], respectively, and the proofs are identical.

We write down the explicit local equations since we will need them for the following lemma. The formal completion of $\tilde{\mathbb{S}}_{K_1}$ at a singular point is isomorphic to

$$(3.3.8.1) \qquad\qquad Spf(S[[\zeta_q, x, y]]/(xy - w)),$$

where $w$ is a uniformizing parameter in $\mathbb{Z}_q[\zeta_q]$.

Let $\omega : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{Z}_q^\times$ be the Teichmüller character.

**LEMMA 3.3.9.** *The diamond operators $<a> \in Aut(\mathfrak{S}_{K_1}(\mathfrak{q})/\mathfrak{S}_{K_0}(\mathfrak{q}))$ extend to automorphisms of $\mathbb{S}_{K_1}$ (resp. $\tilde{\mathbb{S}}_{K_1}$) that act trivially on $(Y_{\acute{e}})_{red}$ (resp. on $\tilde{Y}_m \cap \tilde{Y}_{\acute{e}}$).*

*Proof.* The action of the diamond operators $<a>$ on $\mathfrak{S}_{K_1}(\mathfrak{q})$, in terms of (3.3.6), is deduced from the action on $C^+$ taking a section $c$ to $[a]c$, where $[a]$ denotes multiplication by $a$ (cf. 2.1.4.6). In the model $G_R(u,v)$, with equations (3.3.8) $[a]$ takes $Y$ to $\omega(a)Y$ [TO, p. 13]. The assertion of the lemma follows when $\mathbb{S}_{K_1}$ is replaced by its local model $Spec(R^*)$, in the notation of the proof of Proposition 3.3.7, and this implies the result for $\mathbb{S}_{K_1}$.

In the local model 3.3.8.1 near a singular point of $\tilde{\mathbb{S}}_{K_1}$, $[a]$ acts by taking $x$ to $\omega(a)x$, $y$ to $\omega^{-1}(a)y$ (cf. [DR, V.2.15]). This shows that $[a]$ extends to an automorphism that fixes $\tilde{Y}_m \cap \tilde{Y}_{\acute{e}}$.

## 4. Vanishing cycles and cohomology.

**4.1 Calculations of vanishing cycles.** Let $\eta = Spec(\mathcal{K}_{\mathfrak{q}^{(1)}})$ be the generic point of $S_{\mathfrak{q}}$, $\bar{\eta}$ the spectrum of the algebraic closure of $\mathcal{K}_{\mathfrak{q}^{(1)}}$, $j : \bar{\eta} \to S_{\mathfrak{q}}$ the natural map. We write $Gal(\bar{\eta}/\eta)$ for $Gal(\overline{\mathcal{K}}_{\mathfrak{q}^{(1)}}/\mathcal{K}_{\mathfrak{q}^{(1)}})$, and let $I \subset Gal(\bar{\eta}/\eta)$ denote the inertia subgroup and $P \subset I$ the wild inertia subgroup. The next step is to study the $\ell$-adic cohomology of $\mathfrak{S}_{K_1}(\mathfrak{q})_{\bar{\eta}} = \mathfrak{S}_{K_1}(\mathfrak{q}) \times_\eta \bar{\eta}$ as a module for $Gal(\bar{\eta}/\eta)$, by comparing it to the $\ell$-adic cohomology of the special fiber. It follows from Proposition 3.3.7 that the model $\mathbb{S}_{K_1}$ satisfies hypotheses (i)-(iv) of [RZ1,(2.2)], and that each irreducible component of the special fiber occurs with multiplicity prime to $q$. Then Grothendieck's purity conjecture [SGA 7, I, §3] follows from [RZ1, (2.21)]. It follows that the calculation in [SGA 7, I, 3.3] of the sheaves of tame vanishing cycles is valid.

**4.1.1.** We follow the notation of [RZ1] and let $R\Psi^q\mathbb{Q}_\ell$ denote the $q$th vanishing cycle sheaf. By [RZ1, (2.23)], the vanishing cycle sheaves and the tame vanishing cycle sheaves coincide under our present hypotheses. We recall the calculation of $R\Psi^q\mathbb{Q}_\ell$ in the case at hand, which we will treat abstractly.

Consider a regular scheme $X$, flat and of finite type over $S_{\mathfrak{q}}$, with smooth generic fiber $X_\eta$ and special fiber $X_s$. The special fiber is assumed to be étale locally of one of the following types:

**4.1.1.1.** $X_s$ is smooth. This corresponds to a neighborhood of a point in $Y_m$ not in $Y_{\acute{e}}$. Then the regularity of $X$ implies that $X$ is smooth. In this case $R\Psi^q\mathbb{Q}_\ell = 0$ for $q > 0$, $R\Psi^0\mathbb{Q}_\ell = \mathbb{Q}_\ell$.

**4.1.1.2.** $X$ is of the form $Spec\ \mathbb{Z}_q[X, Y, Y^{-1}]/(X^{q-1}Y - q)$. This corresponds to a neighborhood of a point in $Y_{\acute{e}}$ not in $Y_m$. Let $X' = Spec\ \mathbb{Z}_q[x, y, y^{-1}]/(x^{q-1} - q)$, and define a map $\pi : X' \to X$ via the embedding

$$\mathbb{Z}_q[X, Y, Y^{-1}]/(X^{q-1}Y - q) \to \mathbb{Z}_q[x, y, y^{-1}]/(x^{q-1} - q); \ Y \mapsto y^{q-1}, \ X \mapsto xy^{-1}.$$

The morphism $\pi$ is étale with Galois group $\mathbb{F}_q^\times \cong \mu_{q-1}$, the group of $(q-1)$st roots of unity. Here $\zeta \in \mu_{q-1}$ acts by multiplying both $x$ and $y$ by $\zeta$. The special fiber $X_s$ is a non-reduced divisor with multiplicity $q-1$; the associated reduced divisor $X_{s,red}$ is defined by the equation $X = 0$ and is smooth; it is isomorphic to $Spec\ \mathbb{F}_q[Y, Y^{-1}]$, which we view as $(\mathbb{G}_m)_{\mathbb{F}_q}$.

We first compute the vanishing cycle sheaves $R\Psi^q\mathbb{Q}_\ell$ in the étale neighborhood $X'$ of $X$. Let $\mathcal{O} = \mathbb{Z}_q[\zeta_q] = \mathbb{Z}_q[x]/(x^{q-1} - q)$. Then we may write

$$(4.1.1.2.1) \qquad\qquad X' = Spec\ \mathcal{O} \times_{Spec\ \mathbb{Z}_q} Y,$$

where $Y = Spec\ \mathbb{Z}_q[y, y^{-1}]$. The second factor is smooth over $\mathbb{Z}_q$, hence $(R\Psi^q\mathbb{Q}_\ell)_{X'}$ is the pullback from $R\Psi^q$ for the finite flat morphism $Spec\ \mathcal{O} \to Spec\ \mathbb{Z}_q$. The morphism is of relative dimension zero, hence $R\Psi^q = 0$ for $q > 0$, and it is elementary to see that $R\Psi^0$ is the group algebra $\mathbb{Q}_\ell[\mu_{q-1}]$, with the inertia group of $\mathcal{O}$ over $\mathbb{Z}_q$ acting as $\mu_{q-1}$.

It follows that $R\Psi^q\mathbb{Q}_\ell = 0$ for $q > 0$, and that $R\Psi^0\mathbb{Q}_\ell$ is a lisse $\ell$-adic sheaf of rank $q - 1$ on $X_{s,red}$ that becomes constant over $X'_{s,red}$. Moreover, since $Gal(X'/X)$ acts as the inertia group on the first factor of (4.1.1.2.1), one sees easily that the canonical action of $Gal(X'/X)$ on $(R\Psi^0\mathbb{Q}_\ell)_{X'_{s,red}}$ identifies the latter with the group algebra $\mathbb{Q}_\ell[Gal(X'/X)]$. It follows that

$$(4.1.1.2.2) \qquad\qquad R\Psi^0\mathbb{Q}_\ell \xrightarrow{\sim} \pi_*\mathbb{Q}_\ell.$$

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

$$(4.1.1.2.3) \qquad\qquad R\Psi^0\mathbb{Q}_\ell = \oplus_\chi R\Psi^0\mathbb{Q}_\ell[\chi],$$

the decomposition with respect to characters of the inertia group. We write $L[\chi]$ for the rank one local system $R\Psi^0\mathbb{Q}_\ell[\chi]$. Let $\chi_0$ denote the trivial character. Consider the embedding

$$j : X_{s,red} = (\mathbb{G}_m)_{\mathbb{F}_q} \hookrightarrow \mathbb{A}^1 = Spec\ \mathbb{F}_q[Y]$$

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

$$(4.1.1.2.4) \qquad\qquad R^0j_*L[\chi] = j_!L[\chi]; \ R^qj_*L[\chi] = 0, \ q > 0 \quad (\chi \neq \chi_0);$$

$$(4.1.1.2.5) \qquad\qquad R^0j_*L[\chi_0] = \mathbb{Q}_\ell; \ R^qj_*L[\chi_0] = 0, \ q > 0.$$

More generally, suppose $X = Spec\ R[X,Y,Y^{-1}]/(X^{q-1}Y - q)$, where $R$ is a smooth $\mathbb{Z}_q$-algebra of finite type. Then $X$ is the fiber product

$$X = Spec\ R \times_{Spec\ \mathbb{Z}_q} Spec\ \mathbb{Z}_q[X,Y,Y^{-1}]/(X^{q-1}Y - q),$$

where the first factor is smooth. We define

$$j_X = 1 \times j : X_{s,red} = Spec\ R \times_{Spec\ \mathbb{Z}_q} (\mathbb{G}_m)_{\mathbb{F}_q} \to Spec\ R \times_{Spec\ \mathbb{Z}_q} Spec\ \mathbb{F}_q[Y].$$

Letting $X_2$ denote the second factor above and letting $p_2$ denote the projection $X \to X_2$, we see that

$$(4.1.1.2.6) \qquad R\Psi^q \mathbb{Q}_\ell = p_2^* R\Psi_{X_2}^q \mathbb{Q}_\ell,$$

where $R\Psi_{X_2}$ denotes the vanishing cycle sheaves for the map from $X_2$ to $Spec\ \mathbb{Z}_q$. In particular, $R\Psi^q \mathbb{Q}_\ell$ vanishes for $q > 0$, while $R\Psi^0 \mathbb{Q}_\ell$ breaks up under the action of the inertia subgroup of $Gal(\overline{\mathbb{Q}}_q/\mathbb{Q}_q)$ as the sum of rank one local systems $L[\chi]$: $L[\chi_0] = \mathbb{Q}_\ell$, whereas $L[\chi]$ for non-trivial $\chi$ satisfies the analogue of (4.1.1.2.4):

$$(4.1.1.2.7) \qquad R^0 j_{X,*} L[\chi] = j_{X,!} L[\chi];\ R^q j_{X,*} L[\chi] = 0,\ q > 0 \quad (\chi \neq \chi_0).$$

In other words, $j_{X,!} L[\chi]$ is the intersection complex (pure perverse sheaf) on $Spec\ R \times_{Spec\ \mathbb{Z}_q} Spec\ \mathbb{F}_q[Y]$ associated to the local system $L[\chi]$ on the open subscheme $X_{s,red}$.

**4.1.1.3.** $X$ is of the form $Spec\ R[X,Y]/(X^{q-1}Y - q)$, where $R$ is a smooth $\mathbb{Z}_q$-algebra of finite type. This corresponds to a neighborhood of a point in $Y_\acute{e} \cap Y_m$. We want to calculate the stalks of $R\Psi^q \mathbb{Q}_\ell$ at a geometric point $\overline{x}$ of the singular locus $X_{sing}$ of the special fiber, defined by $X = Y = 0$. In this case we simply quote the result from [SGA 7, loc. cit]: $R\Psi^q \mathbb{Q}_\ell = 0, q > 1$; $(R\Psi^0 \mathbb{Q}_\ell)_{\overline{x}} = \mathbb{Q}_\ell, (R\Psi^1 \mathbb{Q}_\ell)_{\overline{x}} = \mathbb{Q}_\ell(-1)$, with trivial action of the inertia group on $\mathbb{Q}_\ell$, the $(-1)$ denoting Tate twist.

We now apply the above calculations to the cohomology of $\mathbb{S}_{K_1}$. We write $Y_a = Y_\acute{e} \cap Y_m$. Let $i_m : Y_m \to \mathbb{S}_{K_1}$, $i_e : Y_\acute{e} \to \mathbb{S}_{K_1}$, and $i_a : Y_a \to \mathbb{S}_{K_1}$ be the natural maps. Let $Y_\acute{e}^o$ denote the complement of $Y_a$ in $Y_\acute{e}$, and let $j_e : (Y_\acute{e}^o)_{red} \to (Y_\acute{e})_{red}$ be the open immersion. Then

PROPOSITION 4.1.1.4. *The vanishing cycle sheaves $R\Psi^q \mathbb{Q}_\ell$ are calculated as follows. Let $I$ denote the inertia subgroup of $Gal(\overline{\mathbb{Q}}_q/\mathbb{Q}_q)$. Then the action of $I$ on $R\Psi^q \mathbb{Q}_\ell$ factors through the map to $\mu_{q-1}$ given by the action on $\mathbb{Q}[\zeta_q]$. For a character $\chi$ of $\mu_{q-1}$, let $[\chi]$ denote the $\chi$-isotypic component, and let $\chi_0$ denote the trivial character. Then*

*(i) $R\Psi^0 \mathbb{Q}_\ell[\chi_0] = \mathbb{Q}_\ell$.*

*(ii) $R\Psi^1 \mathbb{Q}_\ell = R\Psi^1 \mathbb{Q}_\ell[\chi_0]$ is a rank one local system supported on $Y_a$, locally isomorphic at any point of $Y_a$ to $\mathbb{Q}_\ell(-1)$.*

*(iii) For $\chi \neq \chi_0$, $R\Psi^0 \mathbb{Q}_\ell[\chi]$ is the extension by zero of a rank one lisse sheaf $L[\chi]$ supported on $Y_\acute{e}^o$. Moreover, the natural map $j_{e,!} R\Psi^0 \mathbb{Q}_\ell[\chi] \to Rj_{e,*} R\Psi^0 \mathbb{Q}_\ell[\chi]$ is a quasi-isomorphism.*

*(iv) Finally, $R\Psi^q \mathbb{Q}_\ell = 0$ for $q > 1$.*

*Proof.* Everything follows immediately from (4.1.1.1-3) except the global triviality of $R\Psi^0 \mathbb{Q}_\ell[\chi_0]$. But there is always an injection $\mathbb{Q}_\ell \to R\Psi^0 \mathbb{Q}_\ell[\chi_0]$, so (i) follows from the fact that all stalks of $R\Psi^0 \mathbb{Q}_\ell[\chi_0]$ are one-dimensional.

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

$$\ldots \to H^i(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0\mathbb{Q}_\ell) \to H^i(\mathbb{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell)$$
$$\to H^{i-1}(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^1\mathbb{Q}_\ell) \to H^{i+1}(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0\mathbb{Q}_\ell) \to \ldots.$$

Using (4.1.1.4.ii), we rewrite this
(4.1.2)
$$\ldots \to H^i(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0\mathbb{Q}_\ell) \to H^i(\mathbb{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell) \to H^{i-1}((Y_a)_{red}, R\Psi^1\mathbb{Q}_\ell[\chi_0]) \to \ldots.$$

We deduce from (i) and (iii) of (4.1.1.4) that the first term in turn is calculated by a long exact sequence

(4.1.3)   $$\ldots H^{i-1}((Y_a)_{red}, \mathbb{Q}_\ell) \to H^i(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0\mathbb{Q}_\ell)$$
$$\to H^i(Y_m, \mathbb{Q}_\ell) \oplus H^i((Y_{\acute{e}})_{red}, \mathbb{Q}_\ell) \oplus \bigoplus_{\chi \neq \chi_0} H^i_c(Y^o_{\acute{e}}, L[\chi]) \to \ldots.$$

Here and in (4.1.2) we have replaced $Y_{\acute{e}}$ and $Y_a$ by the associated reduced schemes, since the étale cohomology is insensitive to nilpotents.

The diamond operators act on $\mathbb{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$ as well as on $Y_{\acute{e}}$ and $Y_m$, and thus induce compatible actions on the spaces in the exact sequences (4.1.2) and (4.1.3). These are determined as follows:

LEMMA 4.1.4. *The diamond operators $< a >$ act on the outer terms of the exact sequence (4.1.3) as follows: The action is trivial on $H^\bullet((Y_{\acute{e}})_{red}, \mathbb{Q}_\ell)$, $H^\bullet((Y_a)_{red}, \mathbb{Q}_\ell)$, and on $H^{i-1}((Y_a)_{red}, R\Psi^1\mathbb{Q}_\ell[\chi_0])$ and acts via $\chi$ on $L[\chi]$.*

*Proof.* By Lemma 3.3.10 the diamond operators act trivially on $(Y_{\acute{e}})_{red}$ and $(Y_a)_{red}$, so it suffices to determine their action on $R\Psi^0\mathbb{Q}_\ell$ and on $R\Psi^1\mathbb{Q}_\ell$.

We first determine the action on $R\Psi^0\mathbb{Q}_\ell$. By the calculation in (4.1.1.2), we see it suffices to determine the action of the diamond operators on $H^0(Spec(\overline{\mathbb{Q}}_q \otimes_{\mathbb{Q}_q} \mathbb{Q}_q[\zeta_q], \mathbb{Q}_\ell)$, via the identification of $\mathbb{Q}_q[\zeta_q]$ with the generic fiber of $\mu_q$ and the latter with $C_1^{univ}$ in (4.1.1.1-2). But the diamond operators on $\mu_q$ are tautologically given by the cyclotomic character.

As for the action on $R\Psi^1\mathbb{Q}_\ell$, this is again local. But locally the calculation in (4.1.1.3) shows that $R\Psi^1\mathbb{Q}_\ell$ is a constant sheaf, so the triviality of the action of the diamond operators is clear.

COROLLARY 4.1.5. *Suppose $\chi \neq \chi_0$, and denote by $<>=\chi$ the $\chi$-isotypic component for the action of the diamond operators. Then for any $i$, there is a canonical isomorphism of $Gal(\overline{\eta}/\eta)$-modules*

$$H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi} \oplus H^i_c((Y^o_{\acute{e}})_{red}, L[\chi]) \xrightarrow{\sim} H^i(\mathbb{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell)^{<>=\chi}$$

*Proof.* Indeed, in (4.1.3), the diamond operators act trivially on the term $H^{i-1}((Y_a)_{red}, \mathbb{Q}_\ell)$ and coincide with inertia on $L[\chi]$, inducing an isomorphism

$$H^i(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0\mathbb{Q}_\ell)^{<>=\chi} \xrightarrow{\sim} H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi} \oplus H^i_c(Y^o_{\acute{e}}, L[\chi]).$$

Similarly, the diamond operators act trivially on the $H^{i-1}((Y_a)_{red}, R\Psi^1\mathbb{Q}_\ell[\chi_0])$ term in (4.1.3).

Over $\tilde{S}_{\mathfrak{q}} = Spec(\mathcal{O}_{\mathfrak{q}}[\zeta_q])$ the special fiber of $\tilde{\mathbb{S}}_{K_1}$ is a reduced divisor with normal crossings (Prop. 3.3.8). Let $R\tilde{\Psi}^q\mathbb{Q}_\ell$ denote the corresponding vanishing cycle sheaves. It follows that there is a canonical isomorphism

$$(4.1.6) \qquad\qquad R\tilde{\Psi}^0\mathbb{Q}_\ell \simeq \mathbb{Q}_\ell.$$

The following lemma is immediate from the above calculation.

LEMMA 4.1.7. *We identify $(Y^o_\acute{e})_{red} \simeq X^o_\acute{e}$ via the morphism $p_1$. Then the morphism $\tilde{p}_1 : \tilde{Y}^o_e \to (Y^o_\acute{e})_{red}$ of Proposition 3.3.8 defines a canonical isomorphism*

$$\tilde{p}_{1,*}R\tilde{\Psi}^0\mathbb{Q}_\ell \simeq \tilde{p}_{1,*}\mathbb{Q}_\ell \simeq j^*_e(R\Psi^0\mathbb{Q}_\ell).$$

*Moreover, this isomorphism is equivariant with respect to the diamond operators.*

*Proof.* It is easy to see that the étale covering $X'$ of $X$ used in the calculation of 4.1.1.2 is represented by $Isom_X(\mu_q, C_1)$. It follows from Prop. 3.3.8 (d) that the calculation globalizes to give the indicated isomorphism. The diamond operators act as $Aut(\mu_q) \simeq \mu_{q-1}$ on $X'$, and the final assertion follows from the discussion in 4.1.1.2.

COROLLARY 4.1.8. *Let $\chi \neq \chi_0$ be a non-trivial character of the diamond operators. Then the canonical maps*

$$H^i_c((Y^o_\acute{e})_{red}, L[\chi]) \to H^i_c(\tilde{Y}^o_e, \mathbb{Q}_\ell)^{<>=\chi} \to H^i(\tilde{Y}_\acute{e}, \mathbb{Q}_\ell)^{<>=\chi}$$

*are isomorphisms.*

*Proof.* The first isomorphism is a consequence of the preceding lemma. The second isomorphism follows from the fact that the diamond operators act trivially on the complement $\tilde{Y}_m \cap \tilde{Y}_\acute{e}$ of $\tilde{Y}^o_e$ in $\tilde{Y}_\acute{e}$.

## 4.2. The congruence formula and proof of Theorem 2.3.5.

As explained in §2.2, the double coset $U_\mathfrak{q}$ defines a correspondence $\mathbf{U}_\mathfrak{q}$ on $\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}} \times \mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$. Its modular interpretation in characteristic zero has been described in §2.2. An alternative description will be useful in defining its reduction (mod q). Consider the finite flat group scheme $\mathbf{X}(A)_2[q^2]$ over $\mathfrak{S}_{K_1,Q}$, and let $m_q : \mathbf{X}(A)_2[q^2] \to \mathbf{X}(A)_2[q]$ denote multiplication by $q$. Let $P_\zeta = \gamma^*(\zeta)$, $\zeta \in \mu_q$, as in the discussion in 2.2. The inverse image $m_q^{-1}(P_\zeta) \subset \mathbf{X}(A)_2[q^2]$ defines an étale covering of $\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$, whose image under $\rho$, $\rho(m_q^{-1}(P_\zeta)) \subset \mathbf{X}(A')_2[q]$ is a well-defined étale covering of $\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$. Over $\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}} \times \mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$, $\mathbf{U}_\mathfrak{q}$ is in one-to-one correspondence with the set of

$$(4.2.1) \qquad\qquad \{(A, \iota, \lambda, \beta^q, P; A', \iota', \lambda', (\beta')^q, P')\}$$

with notation as in (2.2.2). Here $P = \gamma(1) \in \mathbf{X}(A')_1[q]$, $P' = \gamma'(1) \in \mathbf{X}((A')')_1[q]$, where $(A')'$ bears the same relation to $A'$ and $Im((\gamma')^*) \in \mathbf{X}(A')_2$ that $A'$ bears to $A$ and $Im(\gamma^*)$. Recall that $P$ and $P'$ determine $P_\zeta = \gamma^*(\zeta)$ and $P'_\zeta = (\gamma')^*(\zeta)$ as in (2.2.1.3), and $(P_\zeta, P'_\zeta) \subset \mathbf{X}(A)_2[q] \times \mathbf{X}(A')_2[q]$ is defined by the closed relation

$$(4.2.2) \qquad\qquad P'_\zeta \in \rho(m_q^{-1}(P_\zeta)).$$

This subset is independent of the choice of $\zeta$ and defines a closed relation on the set of $(P, P') \in \mathbf{X}(A')_1[q] \times \mathbf{X}((A')')_1[q]$.

Let $s : \mathfrak{S}_{K_1,Q} \to \mathbf{X}(A')_1[q]$ denote the canonical section sending $(A, \iota, \lambda, \beta^q, P)$ to $P$. Then (4.2.1) is contained in $s(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}) \times s(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}})$. We use the same notation

$\mathbf{U_q}$ to denote the scheme-theoretic closure of (4.2.1) in $s(\mathbb{S}_{K_1}) \times s(\mathbb{S}_{K_1})$. Let $p_i$ : $\mathbf{U_q} \to \mathbb{S}_{K_1}$ $i = 1, 2$, denote projections on the $i$th factor.

Over the open subset $Y_{\acute{e}}^o$, defined to be the complement of $Y_a$ in $Y_{\acute{e}}$, the morphisms $p_1$ and $p_2$ are both étale. Moreover, it is straightforward to see that $p_2(p_1^{-1}(Y_{\acute{e}}^o)) \subset Y_{\acute{e}}^o$. In particular, the modular description (4.2.1) holds without change over $Y_{\acute{e}}^o$. We let $\mathbf{U_{\acute{e}}}$ denote the restriction of $\mathbf{U_q}$ to $Y_{\acute{e}}^o \times Y_{\acute{e}}^o$.

The situation is more interesting over $Y_m^o$. First, suppose $(A, \iota, \lambda, \beta^q, P)$ is a ($T$-valued) point of $Y_m^o$. Then the étale group $C_1$ generated by $P$ corresponds (by Cartier duality) to a subgroup $C_{2,P}$ of $\mathbf{X}(A)_2$ isomorphic to $\mu_q$. The isogeny $F_2$ : $\mathbf{X}(A)_2 \to \mathbf{X}(A')_2 = \mathbf{X}(A)_2/C_P$ is just the geometric Frobenius map. Indeed, one verifies that $A'$ is nearly isomorphic to the transform $A^{(q)}$ of $A$ under Frobenius. In fact, this is true at the level of q-divisible groups: writing

$$\mathbf{X}(A') = [\mathbf{X}(A')_1 \oplus \mathbf{X}(A')_2]^n,$$

we have canonical isomorphisms

$$(4.2.3) \quad \mathbf{X}(A')_2 \xrightarrow{\sim} \mathbf{X}(A^{(q)})_2; \quad m_q \cdot \mathbf{X}(A')_1 \xrightarrow{\sim} \mathbf{X}(A)_1/r'(\mathbf{X}(A')_1[q]) \xrightarrow{\sim} \mathbf{X}(A^{(q)})_1$$

where the first isomorphism is given by $F_2$, the second is obtained by factorizing $m_q : \mathbf{X}(A')_1 \to \mathbf{X}(A')_1$ via the morphism $r' : \mathbf{X}(A')_1 \to \mathbf{X}(A)_1$ of 2.1.4, and the third is a consequence of our choice of $F_2$. Moreover, the point $P'$ of (4.2.1) satisfies the relation

$$P' \in m_q(r')^{-1}(P)$$

derived from (4.2.2), as one verifies by looking at the multiplicative formal group. Here the isogeny $r' : \mathbf{X}((A')')_1 \to \mathbf{X}(A')_1$ is again as in 2.1.4, but this time relative to $A'$! We note that, identifying $m_q \cdot \mathbf{X}(A')_1$ with $\mathbf{X}(A')_1$, $P'$ can be identified with the image of $P$ under the Frobenius map

$$\mathbf{X}(A)_1 \times \mathbf{X}(A)_2 \to \mathbf{X}(A^{(q)})_1 \times \mathbf{X}(A^{(q)})_2.$$

It follows that

$$(4.2.4) \qquad (A, \iota, \lambda, \beta^q, P) \to \{(A, \iota, \lambda, \beta^q, P; A', \iota', \lambda', (\beta')^q, P')\}$$

is the graph of a morphism $\mathbf{U_m^+} : Y_m^o \to Y_m^o$ that defines a section of $p_1$ over $Y_m^o$. Here the quadruple $(A', \iota', \lambda', (\beta')^q)$ is defined by $A$ and the isogeny $F_2 : \mathbf{X}(A)_2 \to \mathbf{X}(A')_2$ as in 2.1.4.

Next, let $\mathbf{U_m^-} = p_1^{-1}(Y_m^o) \cap p_2^{-1}(Y_{\acute{e}}^o)$. Then we have

$$(4.2.5) \qquad\qquad p_1^{-1}(Y_m^o) = \mathbf{U_m^+} \sqcup \mathbf{U_m^-}.$$

In particular, $\mathbf{U_m^+}$ (resp. $\mathbf{U_m^-}$) is the graph of a correspondence on $Y_m^o \times Y_m^o$ (resp. $Y_m^o \times Y_{\acute{e}}^o$). We denote their closures in the special fiber of $\mathbb{S}_{K_1} \times \mathbb{S}_{K_1}$ by the same notation.

The morphism $\mathbf{U_m^+}$ is not quite equal to geometric Frobenius on $Y_m^o$. Recall from 3.2.1(b) that the q-divisible $\mathcal{O}_{\mathcal{K},q}$-module $\mathbf{X}(A)$ breaks up over the divisors of $q$ in $\mathcal{K}$. The components corresponding to divisors of $q^{(1)}$ (resp. $q^{(2)}$) other than $\mathfrak{q}^{(1)}$ (resp. $\mathfrak{q}^{(2)}$) are of multiplicative type (resp. étale). Thus geometric Frobenius on $Y_m^o$ acts as multiplication by $q$ on the $q^{(1)}$-factors of $\mathbf{X}(A)$ (other than $\mathfrak{q}^{(1)}$) and as the identity on the $q^{(2)}$-factors (other than $\mathfrak{q}^{(2)}$). Moreover, the maximal connected subgroup

$\mathbf{X}(A)_1^0$ is purely of multiplicative type. Thus geometric Frobenius replaces $\mathbf{X}(A)_1$ by $\mathbf{X}(A)_1/\mathbf{X}(A)_1^0[q]$. But it follows easily from (2.1.4.2) that

$$\mathbf{X}(A)_1/\mathbf{X}(A)_1^0[q] = \mathbf{X}(A'')/\mathbf{X}(A'')[q].$$

We combine these observations as follows. Let $[q^{-1}]_1$ denote the image of $q^{-1}$ in $GL(n, \mathcal{K}_{0,q^{(1)}})$, which we embed diagonally in $GL(n, \mathcal{K}_{q^{(1)}})$. This is a central element of $G(\mathbf{A}_f)$, hence induces automorphisms of all Shimura varieties at finite level that extend trivially to $\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}$. Bearing in mind that inclusions of lattices induce maps of abelian varieties in the opposite direction, we conclude that

PROPOSITION 4.2.6 (CONGRUENCE FORMULA). *We have the identity*

$$[q^{-1}]_1 \cdot \mathbf{U}_m^+ = Frob,$$

*the geometric Frobenius map on the scheme $Y_m^o$ over $k(\mathfrak{q}^{(1)})$.*

**4.2.7.** We need to explain how to relate the cohomological correspondence defined by $\mathbf{U}_\mathfrak{q}$ in characteristic zero to the finite correspondences calculated above over $Y_{\acute{e}}^o$ and $Y_m^o$. This is somewhat delicate, since the base scheme $\mathbb{S}_{K_1}$ has a singular special fiber and we have not determined the structure of $\mathbf{U}_\mathfrak{q}$ over the singular locus. Fortunately, we are only interested in the action on $R\Psi^0$. It is convenient to write

$$(4.2.7.1) \qquad R\Psi^0 \mathbb{Q}_\ell = \mathbb{Q}_\ell \oplus \bigoplus_{\chi \neq \chi_0} j_{m,!} L[\chi].$$

We consider the first map of (4.1.2):

$$b_i : H^i(\mathbb{S}_{K_1,\overline{\kappa}}, R\Psi^0 \mathbb{Q}_\ell) \to H^i(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell).$$

We need to find a cohomological correspondence $\mathbf{U}_\mathfrak{q}^0$ on $R\Psi^0 \mathbb{Q}_\ell$ such that

$$(4.2.7.2) \qquad b_i \circ \mathbf{U}_\mathfrak{q}^0 = \mathbf{U}_\mathfrak{q} \circ b_i.$$

We define $\mathbf{U}_\mathfrak{q}^0$ on the separate summands on the right-hand side of (4.2.7.1). On the summand $\mathbb{Q}_\ell$ one just takes the natural cohomological correspondence (pullback followed by push-forward) defined by the proper correspondence $\mathbf{U}_\mathfrak{q}$. Recall from Corollary 4.1.5 that, if $\chi \neq \chi_0$, then

$$(4.2.7.3) \qquad H^i(\mathbb{S}_{K_1,\overline{\kappa}}, \mathbb{Q}_\ell)^{<>=\chi} = H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi}.$$

Now $\mathbf{U}_\mathfrak{q}$ commutes with the diamond operators, hence fixes the $\chi$-eigenspace. We have seen that $\mathbf{U}_\mathfrak{q}$ restricts to a sum of two correspondences $\mathbf{U}_m^+ + \mathbf{U}_m^-$ on $Y_m$, and the restriction of $\mathbf{U}_\mathfrak{q}^0$ to $H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi}$ is evidently the sum of the induced cohomological correspondences, say $\mathbf{U}_m^+[\chi]$ and $\mathbf{U}_m^-[\chi]$.

It remains to define $\mathbf{U}_\mathfrak{q}^0$ on the $L[\chi]$. In fact, it is enough to define $\mathbf{U}_\mathfrak{q}^0$ on $H_c^i((Y_{\acute{e}}^o)_{red}, L[\chi])$. Extending scalars to $\tilde{S}_\mathfrak{q}$, we can use the model $\tilde{\mathbb{S}}_{K_1}$, and define a correspondence $\tilde{\mathbf{U}}_\mathfrak{q}$ on $\tilde{\mathbb{S}}_{K_1} \times \tilde{\mathbb{S}}_{K_1}$ as above. We have seen that $R\tilde{\Psi}^0 \mathbb{Q}_\ell = \mathbb{Q}_\ell$ (4.1.6), hence we can define $\tilde{\mathbf{U}}_\mathfrak{q}^0$ again as the natural cohomological correspondence on $R\tilde{\Psi}^0 \mathbb{Q}_\ell$. This commutes with the action of the diamond operators, hence for each $\chi$ defines an operator on $H^i(\tilde{\mathbb{S}}_{K_1,\overline{\kappa}}, \mathbb{Q}_\ell)^{<>=\chi}$. Since the diamond operators act trivially on $\tilde{Y}_{\acute{e}} \cap \tilde{Y}_m$, it follows as before that, for $\chi \neq \chi_0$, this breaks up as

$$H^i(\tilde{Y}_m, \mathbb{Q}_\ell)^{<>=\chi} \oplus H^i(\tilde{Y}_{\acute{e}}, \mathbb{Q}_\ell)^{<>=\chi} = H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi} \oplus H_c^i((Y_{\acute{e}}^o)_{red}, L[\chi]),$$

by Proposition 3.3.8 (b) and Corollary 4.1.8. It is clear that $\tilde{\mathbf{U}}_{\mathfrak{q}}^0$ induces $\mathbf{U}_m^+[\chi] \oplus \mathbf{U}_m^-[\chi]$ on the first factor. On the second factor it suffices to know that $\tilde{\mathbf{U}}_{\mathfrak{q}}$ fixes $\tilde{Y}_{\acute{e}}$. By proper base change, $\tilde{\mathbf{U}}_{\mathfrak{q}}^0$ satisfies (4.2.7.2). Hence

LEMMA 4.2.8. *Fix an integer $i$. Let $\chi \neq \chi_0$ be a non-trivial character of the diamond operators. The correspondence $\mathbf{U}_{\mathfrak{q}}$ acts on*

$$H^i(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell)^{<>=\chi} \cong H^i(Y_m, \mathbb{Q}_\ell)^{<>=\chi} \oplus H^i_c(Y^o_{\acute{e}}, L[\chi])$$

*via the matrix*

$$\begin{pmatrix} [q^{-1}]_1^{-1} \cdot Frob & 0 \\ 0 & D(\chi) \end{pmatrix}$$

*where $[q^{-1}]_1$ is the automorphism defined above, $Frob$ is geometric Frobenius, lifted to $\mathcal{K}_{\mathfrak{q}}[\zeta_q]$, and $D(\chi) \in Aut(H^i(\tilde{Y}_{\acute{e}}, \mathbb{Q}_\ell)^{<>=\chi})$.*

*Proof.* Everything is clear except the $0$ in the upper right-hand corner. But it follows from Lemma 4.1.4 and Proposition 4.1.1 that the inertia group $I$ acts trivially on $H^i(Y_m, \mathbb{Q}_\ell)$ and coincides with the diamond operators on $H^i(Y^o_{\acute{e}}, L[\chi])$. The correspondence $\mathbf{U}_{\mathfrak{q}}$ is defined over $S_{\mathfrak{q}}$, hence commutes with inertia. The Proposition follows.

### 4.2.9.

*Proof of Theorem 2.3.5.* We can now complete the proof of Theorem 2.3.5. We are given an automorphic representation $\pi$ that contributes to $H^{n-1}(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell)$ but has no $K_0(\mathfrak{q})$-fixed vector; i.e., we are in case (b) of Proposition 1.4. Let $\{\alpha, \beta_1, \ldots, \beta_{n-1}\}$ be the characters defined there, so that $\alpha$ is tamely ramified and the $\beta_i$ are unramified. Let $\chi$ denote the restriction of $\alpha$ to the inertia subgroup; thus $\chi$ coincides with the action of $K_0(\mathfrak{q})/K_1(\mathfrak{q})$ on the $K_1(\mathfrak{q})$-fixed vectors in $\pi$. In particular, and bearing in mind the discussion in 2.1.4.6, $\pi$ contributes to $H^{n-1}(\mathfrak{S}_{K_1}(\mathfrak{q})_{\overline{\eta}}, \mathbb{Q}_\ell)^{<>=\chi^{-1}}$. Define $M_\lambda[\pi_f]$ as in §2.3. The two components $Y_m$ and $Y_{\acute{e}}$ are invariant under prime-to-$q$ isogenies, hence $M_\lambda[\pi_f]$ breaks up as above as the sum of $A^\# = H^{n-1}(Y_m, \mathbb{Q}_\ell)^{<>=\chi^{-1}}[\pi_f]$ and $B^\# = H^{n-1}_c(Y^o_{\acute{e}}, L[\chi^{-1}])[\pi_f]$. Inertia at $\mathfrak{q}$ acts on the first factor trivially and on the second factor by the character $\chi^{-1}$. Then

$$r_\rho(\pi) \xrightarrow{\sim} A^\# \otimes \alpha(\pi) \oplus B^\# \otimes \alpha(\pi) = A \oplus B.$$

But

$$\alpha(\pi) = \xi_\pi \circ N_{\mathcal{K}/\mathcal{K}_0} = \alpha \cdot \prod_i \beta_i.$$

On the one hand

$$\alpha(\pi)(\varpi) = \xi_\pi(N_{\mathcal{K}/\mathcal{K}_0}(\varpi)) = \xi_\pi([q^{-1}]_1)^{-1}$$

On the other hand, the restriction to inertia of $\xi_\pi$ is the same as that of $\alpha$, namely $\chi$. It follows from Lemma 4.2.8 that

**4.2.9.1.** Geometric Frobenius coincides with $\mathbf{U}_{\mathfrak{q}}$ on $A$. Moreover, inertia acts as $\chi$ on $A$.

**4.2.9.2.** The representation at $\mathfrak{q}^{(1)}$ on $B$ is unramified.
These two assertions complete the proof of Theorem 2.3.5.

**4.2.10.** In particular, inertia at $\mathbf{q}^{(1)}$ acts on the determinant of $r_\rho(\pi)$ by $\chi^{\dim A}$. But the determinant of $r_\rho(\pi)$ is given by $\xi_\pi$ at all unramified places, hence at $\mathbf{q}^{(1)}$ by Chebotarev density. It follows that

$$(4.2.10.1) \qquad\qquad \chi^{\dim A - 1} = \chi_0.$$

Now suppose

$$(4.2.10.2) \qquad\qquad q \equiv 1 \pmod{\ell}, \; \ell > n, \; \chi|_{\Delta_q} \neq 1,$$

where $\Delta_q$ denotes the $\ell$-Sylow subgroup of $\mathbb{F}_q^\times$. It follows from (4.2.10.1) that $\ell | (\dim A - 1)$. Under hypotheses (4.2.10.2), this is only possible if $\dim A = 1$. Thus

COROLLARY 4.2.10.3. *Under the hypotheses (4.2.10.2), the spaces $A$ and $B$ in the statement of Theorem 2.3.5 are of dimension 1 and $n - 1$, respectively.*

## REFERENCES

[AC] ARTHUR, J. AND L. CLOZEL, *Simple Algebras, Base Change, and the Advanced Theory of the Trace Formula*, Annals of Math Studies, 120 (1989).

[BZ] BERNSTEIN J. AND A. ZELEVINSKI, *Induced representations of reductive p-adic groups I.*, Ann. Sci. Ec. Norm. Sup., 10 (1977), pp. 441–472.

[Ca] CARAYOL, H., *Sur la mauvaise réduction des courbes de Shimura*, Compositio Math., 59 (1986), pp. 151–230.

[C1] CLOZEL, L., *Représentations Galoisiennes associées aux représentations automorphes autoduales de GL(n)*, Publ. Math. I.H.E.S., 73 (1991), pp. 97–145.

[C2] CLOZEL, L., *On the cohomology of Kottwitz's arithmetic varieties*, Duke Math. J., 72 (1993), pp. 757–795.

[CL] CLOZEL, L. ET J.-P. LABESSE, *Changement de base pour les représentations cohomologiques de certains groupes unitaires, appendix to [L]*.

[dJ] DE JONG, J., *The moduli spaces of principally polarized abelian varieties with $\Gamma_0(p)$-level structure*, J. Alg. Geom., 2 (1993), pp. 667–688.

[DR] DELIGNE, P. AND M. RAPOPORT, *Les schémas de modules de courbes elliptiques, in Modular Functions of One Variable II*, Lect. Notes in Math., 349 (1973), pp. 143–316.

[G] GÖRTZ, U., *On the flatness of models of certain Shimura varieties of PEL type*, Math. Ann., 321 (2001), pp. 689-727.

[Gr] GROSS, B., *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math. J., 61 (1990), pp. 445–517.

[H1] HARRIS, M., *Supercuspidal representations in the cohomology of Drinfel'd upper half spaces; elaboration of Carayol's program*, Invent. Math., 129 (1997), pp. 75–119.

[H2] HARRIS, M., *The local Langlands conjecture for GL(n) of a p-adic field, $n < p$*, Invent. Math., 134 (1998), pp. 177–210.

[HL] HARRIS, M. AND J.-P. LABESSE, *Conditional base change for unitary groups*, (in preparation).

[HT1] HARRIS, M. AND R. TAYLOR, *Deformations of automorphic Galois representations*, (preprint, 1998-99).

[HT2] HARRIS, M. AND R. TAYLOR, *On the geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, Princeton: Princeton University Press, 151 (2001).

[He] HENNIART, G., *Une preuve simple des conjectures de Langlands pour $GL_n$ sur un corps p-adique*, Invent. Math., 139 (2000), pp. 439–455.

[HH] HENNIART, G. AND R. HERB, *Automorphic induction for GL(n) over local non-archimedean fields*, Duke. Math. J., 78 (1995), pp. 131–192.

[JPSS] JACQUET, H., I. I. PIATETSKI-SHAPIRO, AND J. SHALIKA, *Conducteur des représentations génériques du groupe linéaire,*, C.R.Acad. Sc. Paris, 292 (1981), pp. 611–616.

[KM] KATZ, N.M. AND B. MAZUR, *Arithmetic Moduli of Elliptic Curves*, Annals of Math. Studies, 108 (1985 Princeton: Princeton University Press).

[K1] KOTTWITZ, R., *On the λ-adic representations associated to some simple Shimura varieties*, Invent. Math., 108 (1992), pp. 653–665.

[K2] KOTTWITZ, R., *Points on some Shimura varieties over finite fields*, J. Am. Math. Soc., 5 (1992), pp. 373–444.

[L] LABESSE, J.-P., *Cohomologie, stabilisation et changement de base (avec des appendices par Clozel-Labesse et Breen)*, Astérisque, 257 (1999).

[Me] MESSING, W., *The crystals associated to Barsotti-Tate groups with applications to abelian schemes*, Lect. Notes Math., 264 (1972).

[R] RAPOPORT, M., *On the bad reduction of Shimura varieties*, in L. Clozel and J.S. Milne, eds., *Automorphic Forms, Shimura Varieties, and L-functions*, New York: Academic Press, Vol. II (1990), pp. 253–321.

[RZ1] RAPOPORT, M. AND T. ZINK, *Uber die lokale Zetafunktion von Shimuravarietäten Monodromiefiltration und verschwindende Zyklen in ungleicher Characteristik*, Invent. Math., 68 (1982), pp. 21–101.

[RZ2] RAPOPORT, M. AND T. ZINK, *Period spaces for p-divisible groups*, Annals of Math. Studies, 141 (1996).

[SGA7] , *Groupes de Monodromie en géométrie Algébrique I*, Lect. Notes in Math., 288 (1972).

[SW] SKINNER, C. AND A WILES, *Base change and a problem of Serre*, Duke Math. J., 107 (2001), pp. 15–25.

[TO] TATE, J. AND F. OORT, *Group schemes of prime order*, Ann. Sci. Ec. Norm. Sup., 3 (1970), pp. 1–21.

[TW] TAYLOR, R. AND A. WILES, *Ring-theoretic properties of certain Hecke algebras*, Annals of Math., 141 (1995), pp. 553–572.

[W] WILES, A., *Modular elliptic curves and Fermat's last theorem*, Annals of Math., 142 (1995), pp. 443–551.

[We1] WEDHORN, T., *Ordinariness in good reductions of Shimura varieties of PEL-type*, Ann. Sci. de l'ENS, 32 (1999), pp. 575–618.

[We2] WEDHORN, T., *Congruence relations on some Shimura varieties*, J. Reine Angew. Math., 524 (2000), pp. 43–71.