J. ALGEBRAIC GEOMETRY
9 (2000) 111–154

STRATIFICATIONS OF HILBERT MODULAR VARIETIES

E. Z. GOREN AND F. OORT

Abstract

We consider g-dimensional abelian varieties in characteristic p, with a given action of $\mathcal{O}_L$, the ring of integers of a totally real field L of degree g. A stratification of the associated moduli spaces is defined by considering the action of $\mathcal{O}_L$ on a certain canonical subgroup of abelian varieties. The properties of this stratification and its relation to other stratifications, e.g. by Newton polygons, are studied.

1. Introduction

There are two important stratifications of the moduli space $\mathcal{A} = \mathcal{A}_{g,1} \otimes \mathbb{F}_p$ of principally polarized abelian varieties of dimension g in characteristic p. One is given by Newton polygons. The other is given by the structure of the p-torsion subgroup scheme. We will call the latter the Ekedahl-Oort stratification, see [EO]. Similar constructions can be made for other Shimura varieties of Hodge type. In this paper we consider a certain stratification of Hilbert modular varieties in positive characteristic. It is a refinement of the Ekedahl-Oort stratification on these varieties. We will examine its relation to the Newton polygon stratification.

The stratification is obtained as follows. Fix a totally real field L of degree g over $\mathbb{Q}$. Let $\mathcal{O}_L$ be the ring of integers of L and let p be a rational prime that is inert in L. Fix an integer $n \geq 3$ and prime to p. Denote by $\mathbb{F}$ a finite field obtained from a field of p^g elements by adjoining a primitive n-th root of unity.

We consider quadruples $\mathbf{X} = (X, \lambda, \iota, \alpha)$ over a field k of characteristic p containing $\mathbb{F}$, consisting of an abelian variety X over k, a principal polarization λ of X, an embedding ι of $\mathcal{O}_L$ into the endomorphisms of X fixed by the Rosati involution associated to λ, and a full symplectic level-n structure α. We shall

Received May 22, 1998 and, in revised form, November 11, 1998.

denote the moduli space of such data $\mathbf{X}$ by $\mathcal{M}_n$. It is a regular, irreducible, g-dimensional variety over $\mathbb{F}$. It is a fine moduli scheme.

Given such $\mathbf{X}$, the cotangent space of X is a free $(\mathcal{O}_L \otimes k)$-module of rank 1. Let us denote the embeddings of $\mathcal{O}_L/p$ into $\mathbb{F}$ by $\{\sigma_1, \cdots, \sigma_g\}$, ordered in such a way that the Frobenius composed with σ_i is σ_{i+1}. Let V denote the Verschiebung map. Consider

$$\ker(V : H^0(\Omega^1_X) \to H^0(\Omega^1_X)).$$

It is a k-vector space and an $\mathcal{O}_L$-module on which $\mathcal{O}_L/p$ acts by a set of characters τ, each with multiplicity one. We call τ the *type* of $\mathbf{X}$ and denote it by $\tau(\mathbf{X})$. The stratification is obtained by the type.

Our main results are as follows: For every set $\tau \subseteq \{\sigma_1, \cdots, \sigma_g\}$ of characters there exists a closed subscheme W_τ of $\mathcal{M}_n$, which is universal with respect to the property $\tau(\mathbf{X}) \supseteq \tau$. We establish the following theorems.

1. W_τ is locally irreducible and locally linear.
2. The generic point of every component of W_τ has type τ.
3. $\dim(W_\tau) = g - |\tau|$.
4. $W_\tau \cap W_\sigma = W_{\tau\cup\sigma}$.
5. The variety W^0_τ, defined as $W_\tau \setminus \bigcup_{\sigma \supsetneqq \tau} W_\sigma$, is quasi-affine.
6. The components of all the W_τ's have a natural structure of a simplicial complex.
7. Let $\mathbb{E}$ be the Hodge bundle over $\mathcal{M}_n$, and let $W_a = \bigcup_{|\tau|=a} W_\tau$. We obtain formulae expressing W_a, considered as an element of the Chow ring, as a combination of certain tautological Chern classes. In particular,

$$W_1 = (p-1)c_1(\mathbb{E}).$$

8. Let $\tau(\mathbf{X}) = \tau(\mathbf{Y})$; then $X[p] \cong Y[p]$ as principally polarized group schemes with $\mathcal{O}_L$-structure.
9. The possible p-divisible groups up to isogeny are $G_{\ell/g} + G_{(g-\ell)/g}$ for $0 \leq \ell \leq g$, and $g \cdot G_{1/2}$. Let $\beta_0 < \beta_1 < \cdots$ be the possible Newton polygons.
10. We say that a type is spaced if it contains no two consecutive σ_i's. Let $\lambda(\tau) = \max\{|\tau'| : \tau' \subseteq \tau, \tau' \text{ is spaced}\}$ (except for g odd and $\tau = \{1, \cdots, g\}$, where we put $\lambda(\tau) = (g+1)/2$). The generic point of every component of W_τ has Newton polygon equal to $\beta_{\lambda(\tau)}$.
11. We define $\mathcal{N}_i$ to be the closed subscheme where the Newton polygon is weakly above β_i. We conjecture that $\dim(\mathcal{N}_i)$ is equal to $g-i$, and prove this in special cases (leaving a more systematic discussion for a future paper). Those cases imply in particular the "weak Grothendieck conjecture" for principally polarized p-divisible groups with real multiplication.

We remark that as far as local computations are concerned, as long as p is unramified, the assumption that p is inert in L is completely technical. We hope to discuss the general case in a future paper.

The results of this paper seem fundamental to the study of Hilbert modular forms mod p and p-adic Hilbert modular forms from a geometric point of view. See [G].

Acknowledgments

The first named author thanks Utrecht University for its hospitality during a visit in autumn 1997, and thanks the second named author for the invitation. The visit, during which this work was done, was part of the 1997-98 special project on "Algebraic curves and Riemann surfaces, geometry and arithmetic" funded by the NWO (the Dutch organization of pure research).

2. Local structure

2.1. Basic definitions. Fix a totally real field L of degree g over $\mathbb{Q}$. Let $\mathfrak{d}_L$ be the different ideal of L. Let $\mathcal{O}_L$ be the ring of integers of L and let p be a fixed rational prime that is inert in L. Fix an integer n greater than or equal to 3 and prime to p. Denote by $\mathbb{F}$ a finite field obtained from a field of p^g elements by adjoining a primitive n-th root of unity.

For any ring R of characteristic p, let $W(R)$ denote the ring of infinite Witt vectors with values in R, and let

$$\sigma : W(R) \to W(R), \qquad \tau : W(R) \to W(R),$$

be, respectively, the ring homomorphism Frobenius and the additive map Verschiebung. Write

$$\mathrm{Emb}(\mathcal{O}_L, W(\mathbb{F})) = \{\sigma_1, \cdots, \sigma_g\},$$

where $\sigma \circ \sigma_i = \sigma_{i+1}$. We abuse notation and write also

$$\mathrm{Emb}(\mathcal{O}_L/p, \mathbb{F}) = \{\sigma_1, \cdots, \sigma_g\}.$$

We shall usually denote abelian varieties by the letters X, Y. We denote by $X[p]$ the p-torsion subgroup scheme of X and by $X(p)$ the p-divisible group. We use X^t to denote the dual abelian variety.

Let $f(X)$ and $a(X)$ be the usual f-number and a-number of X. That is, for X defined over a field k of characteristic p,

$$p^{f(X)} = |X[p](\bar{k})|, \quad a(X) = \dim_k\left(\mathrm{Hom}_k(\alpha_p, X[p])\right).$$

We denote by $\alpha(X)$ the maximal α_p-elementary subgroup of X. The order of $\alpha(X)$ is $p^{a(X)}$.

We denote by $\mathcal{D}(X(p))$ the contravariant Dieudonné module of $X(p)$ and by $\mathbb{D}(X(p))$ the covariant Dieudonné module of $X(p)$. The same notation applies to a finite commutative p-torsion group scheme G.

Let $\mathcal{M}_n$ be the moduli space of quadruples $\mathbf{X} = (X, \lambda, \iota, \alpha)/S$ consisting of:

- a scheme S over $\mathbb{F}$;
- a principally polarized abelian scheme (X, λ) over S;
- an embedding of rings

$$\mathcal{O}_L \hookrightarrow \mathrm{End}(X)^\lambda,$$

where $\mathrm{End}(X)^\lambda$ denotes the endomorphisms of X/S, as an abelian scheme, that are fixed by the Rosati involution associated to λ;

- an isomorphism of symplectic $\mathcal{O}_L$-modules

$$\alpha : \mathcal{O}_L/n\mathcal{O}_L \oplus \mathfrak{d}_L^{-1}/n\mathfrak{d}_L^{-1} \longrightarrow X[n](S).$$

This implies that $X[n]/S$ is a constant group scheme over S, and the last isomorphism can be interpreted as an isomorphism of constant group schemes over S.

We consider $\mathcal{M}_n$ as a scheme over $\mathbb{F}$. It is known to be a fine moduli scheme. It is a regular, irreducible, g-dimensional variety over $\mathbb{F}$. We denote by $\mathcal{M}$ the coarse moduli scheme where we take no level structure—that is, the coarse moduli scheme of triples (X, λ, ι).

We remark that the level structure α plays no significant role beyond enabling us to identify the universal deformation ring of $\mathbf{X}$ with the completion of the local ring of the corresponding point in $\mathcal{M}_n$.

Given a point $t \in \mathcal{M}_n(S)$, we let $\mathbf{X}_t = (X_t, \lambda_t, \iota_t, \alpha_t)$ denote the corresponding object obtained by pulling back the universal object over $\mathcal{M}_n$. We remark that by [DP], Corollaire 2.9, $H^0(\Omega^1_{X/S})$ is a locally free $\mathcal{O}_L/p \otimes \mathcal{O}_S$-module of rank 1.

Definition 2.1.1. Let $\mathbf{X} = (X, \lambda, \iota, \alpha)/k$ be as above with k an algebraically closed field. We define the type of $\mathbf{X}$, $\tau(\mathbf{X}) \subseteq \mathbb{Z}/g\mathbb{Z}$, as follows: the action of $\mathcal{O}_L/p$ on $\mathcal{D}(\alpha(X))$ is given by $a(X)$ different characters

$\{\sigma_{i_1}, \cdots, \sigma_{i_{a(X)}}\}$. We let

$$\tau(\mathbf{X}) = \{i_1, \cdots, i_{a(X)}\}.$$

Notation. Given k matrices $B_1, \cdots, B_k$ of size $\ell \times \ell$ and some i such that $1 \leq i \leq k$, we denote by

$$\mathfrak{d}_i(B_1, \cdots, B_k)$$

the matrix A in $M_k(M_\ell)$ that is everywhere zero except for

$$A_{i,1} = B_1,\ A_{i+1,2} = B_2, \cdots,\ A_{i-1,k} = B_k.$$

For example, for $i = 1$ we get a diagonal matrix with the matrices $B_1, \cdots, B_k$ on the diagonal.

2.2. Review of equi-characteristic deformation theory. We quickly review some of the definitions and results of equi-characteristic deformation theory of abelian varieties and p-divisible groups. Our main references are [NO], [CN], [O3] and [Z]. Further references are cited there.

From abelian varieties to p-divisible groups.

Definition 2.2.1. Let S be a scheme. A p-divisible group of height h (synonym: a Barsotti-Tate group of height h) over S is an inductive system of finite flat commutative group schemes over S,

$$0 = G_0 \subsetneq G_1 \subsetneq \cdots \subsetneq G_n \subsetneq \cdots$$

(closed immersions), such that G_i is of order p^{hi} and

$$G_i = \mathrm{Ker}(p^i : G_{i+1} \to G_{i+1}).$$

Given such a p-divisible group G, one can construct its Serre dual G^t; see [O2], Section 7.6. One defines the codimension of G as the dimension of G^t. We have

$$\dim(G) + \mathrm{codim}(G) = \mathrm{height}(G).$$

Example 2.2.2. Let $X \to S$ be an abelian scheme of relative dimension g; then

$$X(p) \stackrel{\mathrm{def}}{=} \varinjlim_i X[p^i]$$

is a p-divisible group of height $2g$ and dimension g.

Theorem 2.2.3 (Serre-Tate). *Let S be a scheme such that p is locally nilpotent in $\mathcal{O}_S$. Let $\mathcal{I} \subseteq \mathcal{O}_S$ be a nilpotent ideal, $S_0 = \mathrm{Spec}(\mathcal{O}_S/\mathcal{I})$, and $X_0 \to S_0$ be an abelian scheme. The functor*

$$\left\{ \begin{matrix} X \to S \text{ abelian scheme} \\ \text{with } f : X \times_S S_0 \xrightarrow{\sim} X_0 \end{matrix} \right\} \rightsquigarrow \left\{ \begin{matrix} G \to S \ p\text{-divisible group} \\ \text{with } \phi : G \times_S S_0 \xrightarrow{\sim} X_0(p) \end{matrix} \right\} \tag{2.1}$$

given by $X \mapsto X(p)$ is an equivalence of categories.

We refer to [K2], Section 1.2.1, and [Me], Section V.2.3., for more details regarding this theorem.

Remark 2.2.4. A refined statement of Theorem 2.2.3, taking into account polarizations and endomorphisms, is valid. Its formulation is clear.

We will usually apply this for the case where S is a characteristic p scheme, i.e., $p = 0$ on S, to study *equi-characteristic* deformations.

From p-divisible groups to Dieudonné modules.

Let R be a commutative ring with 1 having characteristic p. At the moment R is arbitrary. Later on we will be interested, in particular, in the case where R is a quotient of $k[[t_1, \cdots, t_g]]$ by a prime ideal, where k is a field of characteristic p.

Consider the noncommutative Cartier ring $\mathcal{C}(R) = \mathrm{Cart}_p(R)$ as in [NO], p. 416. For R a perfect ring

$$\mathcal{C}(R) = W(R)[F][[V]],$$

where $W(R)$ is the Witt ring over R and F and V satisfy the following relations:

1. $FV = VF = p$;
2. $F\alpha = \alpha^\sigma F, \ \ V\alpha^\sigma = \alpha V, \ \ V\alpha F = \alpha^\tau, \ \ \forall \alpha \in W(R)$.

Definition 2.2.5. A Dieudonné module M over $\mathcal{C}(R)$ is a left $\mathcal{C}(R)$-module that satisfies the following:

1. V is injective on M and M/VM is a free R-module;
2. M is complete and separated in the V-adic topology.

We recall that a commutative formal group γ over R is a group functor

$$\left\{ \begin{matrix} R\text{-algebras of finite} \\ \text{length as } R\text{-modules} \end{matrix} \right\} \rightsquigarrow \{\text{commutative groups}\},$$

which is represented by a formal R-scheme Γ. Moreover, γ is formally smooth if for every such R-algebra R_1 and for every ideal I_1 of R_1, such that I_1^2 is the zero ideal, the natural map from $\gamma(R_1)$ to $\gamma(R_1/I_1)$ is surjective.

Example 2.2.6. Let $X \to \mathrm{Spec}(R)$ be an abelian scheme. The connected component of the p-divisible group $X(p)$ is a commutative, formally smooth, formal group.

Theorem 2.2.7. *There is a covariant functor,*

$$\left\{\begin{matrix} \text{commutative, formally smooth,} \\ \text{formal groups over } R \end{matrix}\right\} \rightsquigarrow \{\text{Dieudonné } \mathcal{C}(R)\text{-modules}\},$$

which is an equivalence of categories, given by $\Gamma \mapsto \mathcal{C}(\Gamma)$. *Moreover,* Γ *is finite dimensional over* R *if and only if* $\mathcal{C}(\Gamma)$ *is finitely generated over* $\mathcal{C}(R)$. *The tangent space of* Γ *is canonically isomorphic with* $\mathcal{C}(\Gamma)/V \cdot \mathcal{C}(\Gamma)$.

Consider a class of $\mathcal{C}(R)$-modules M given by dividing the free object

$$\bigoplus_{i=1}^{n+d} \mathcal{C}(R)e_i \tag{2.2}$$

by the relations:

1. $Fe_j = \sum_{i=1}^{n+d} \alpha_{ij}e_i, \quad j = 1, \cdots, d,$
2. $e_j = V(\sum_{i=1}^{n+d} \alpha_{ij}e_i), \ \ j = d+1, \cdots, n+d,$

for some $\alpha_{ij} \in W(R) \subseteq \mathcal{C}(R)$.

One can prove that under certain conditions, e.g., R perfect and V topologically nilpotent, this defines a Dieudonné $\mathcal{C}(R)$-module. Any Dieudonné $\mathcal{C}(R)$-module M, where R is a local artinian ring or a complete local noetherian ring, whose reduction is the Dieudonné module of a p-divisible group, can be written in this form ([N], Section 1). One says it is *displayed.* In this case (α_{ij}) is invertible. If R is perfect, M is generated as a $W(R)$-module by $e_1, \cdots, e_{2g}$, and if R is a perfect field this is a basis for M as a $W(R)$-module. Using the notation

$$(\alpha_{ij}) = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

one finds that the Frobenius is given by

$$\begin{pmatrix} A & pB \\ C & pD \end{pmatrix}.$$

The module associated to G^t is given by the basis $f_1, \cdots, f_{n+d}$, with relations

1. $f_j = V(\sum_{i=1}^{n+d} \delta_{ij}f_i), \quad j = 1, \cdots, d,$
2. $Ff_j = \sum_{i=1}^{n+d} \delta_{ij}f_i, \quad j = d+1, \cdots, n+d.$

We have the relation

$$(\delta_{ij}) =: {}^t(\alpha_{ij})^{-1}$$

(see [N], p. 504).

A principal quasi-polarization

$$\lambda : G \to G^t$$

of a p-divisible group G (i.e., $\lambda = \lambda^t$ and λ is an isomorphism), is equivalent to an isomorphism

$$\beta : \mathcal{C}(G) \to \mathcal{C}(G^t),$$

with the same properties.

Studying deformations of Dieudonné modules.

From [N], Theorem 1, and the methods of [NO], Section 1, one can deduce the following (see also [O3], Section 2):

Let k be a perfect field of characteristic p. Let G be a p-divisible group over k of height h and dimension d. Let

$$R = k[[t_{ij} : 1 \leq i \leq d,\ 1 \leq j \leq h - d]].$$

Let

$$(\alpha_{ij}) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

be a matrix giving a Dieudonné module $\mathbb{D}$ in displayed form over $\mathcal{C}(R)$, as in (2.2), where $n = h - d$. Let T_{ij} denote the Teichmüller lift of t_{ij} to the Witt ring $W(R)$ and let $T = (T_{ij})$. The matrix

$$\begin{pmatrix} A + TC & B + TD \\ C & D \end{pmatrix} \tag{2.3}$$

is giving a display for the universal deformation of G over $\mathrm{Spec}(R)$.

Suppose further that λ is a principal quasi-polarization, thus $n = d$, and that $e_1, \cdots, e_{n+d}$ are chosen as a symplectic basis. That is, for $i < j$, we have $\langle e_i, e_j \rangle = 0$ unless j equals $i + d$, and then $\langle e_i, e_j \rangle = 1$. In this case, the universal deformation with principal polarization is given by dividing by the ideal generated by $t_{ij} - t_{ji}$ for all i, j.

Remark 2.2.8. Suppose that $\mathbb{D}$ is the Dieudonné module of the p-divisible group of a principally polarized g-dimensional abelian variety (X, λ). Let $\mathcal{D}$ be the contravariant Dieudonné module. We have the well-known diagram

$$\begin{array}{ccccccccc}
0 & \to & H^0(\Omega^1_X) & \to & H^1_{dR}(X) & \to & H^1(\mathcal{O}_X) & \to & 0. \\
 & & \| & & \| & & \| & & \\
 & & V\mathcal{D}/p\mathcal{D} & & \mathcal{D}/p\mathcal{D} & & \mathcal{D}/V\mathcal{D} & & \\
 & & \| & & & & \| & & \\
 & & \mathrm{Lie}(X)^* & & & & \mathrm{Lie}(X^t) & &
\end{array}$$

The Hasse-Witt matrix is a matrix describing the action of the Frobenius on $H^1(\mathcal{O}_X)$. It is dual to the matrix of the Verschiebung acting on $H^0(\Omega^1_{X^t})$ in the diagram

$$\begin{array}{ccccccccc} 0 & \to & H^0(\Omega^1_{X^t}) & \to & H_1^{dR}(X) & \to & H^1(\mathcal{O}_{X^t}) & \to & 0, \\ & & \| & & \| & & \| & & \\ & & V\mathbb{D}/p\mathbb{D} & & \mathbb{D}/p\mathbb{D} & & \mathbb{D}/V\mathbb{D} & & \\ & & \| & & & & \| & & \\ & & \mathrm{Lie}(X^t)^* & & & & \mathrm{Lie}(X) & & \end{array}$$

hence equal (for the right choice of bases) to the matrix of the Frobenius on the term $\mathrm{Lie}(X)$. Note that when writing the module in a displayed form, $e_{g+1}, \cdots, e_{2g}$ span $H^0(\Omega^1_{X^t})$ and $e_1, \cdots, e_g$ project to a basis of $\mathrm{Lie}(X)$. Therefore:

The matrix $A + TC$ appearing in (2.3) – read mod p – is the Hasse-Witt matrix of the deformation corresponding to T.

Remark 2.2.9. The theory of displays was originally developed by D. Mumford (unpublished). It was expounded and developed by P. Norman and Norman – Oort in [N] and [NO]. It was further studied by T. Zink in a greater generality, applicable for non-equi-characteristic situations (see, e.g., [Z], Introduction, Theorem 9).

The theory developed by Zink is more suitable for treating the case of non-perfect base rings R. We refer the interested reader to [Z], Introduction, Definition 1.1, Equation 1.6, Definition 1.6, Example 1.9, Proposition 1.10 (and the following remark), Formula 2.31 and Section 2.2, for a quick overview of how the theories are tied together. We remark that the computations are virtually the same due to the fact that both objects are defined by the "same" equations (see [Z], Equation 1.6).

Below, the computations done for $R = k[[t_1, \cdots, t_g]]$ and similar rings can be interpreted as taking place in its perfect closure, or in terms of displays in the sense of [Z].

2.3. The type and local deformations. Let $t \in \mathcal{M}_n(k)$ be a geometric point. Let $\mathbb{D}$ be the covariant Dieudonné module of X_t with the induced $\mathcal{O}_L$-structure and perfect alternating pairing $\langle \cdot, \cdot \rangle$. We have a decomposition

$$\mathcal{O}_L \otimes W(k) = \bigoplus_{i \in \mathbb{Z}/g\mathbb{Z}} W(k), \tag{2.4}$$

given by

$$a \otimes 1 \mapsto (\sigma_1(a), \cdots, \sigma_g(a)).$$

The ring $\mathcal{O}_L \otimes W(k)$ acts on $\mathbb{D}$ and induces a decomposition

$$\mathbb{D} = \bigoplus_{i \in \mathbb{Z}/g\mathbb{Z}} \mathbb{D}_i. \tag{2.5}$$

The following lemma is essentially known (see [Ra], Proposition 3.5). We include the proof for completeness.

Lemma 2.3.1. *The decomposition* (2.5) *has the following properties:*
(i) $F(\mathbb{D}_i) \subseteq \mathbb{D}_{i+1}$, $\dim_k(\mathbb{D}_{i+1}/F(\mathbb{D}_i)) = 1$;
(ii) $V(\mathbb{D}_i) \subseteq \mathbb{D}_{i-1}$, $\dim_k(\mathbb{D}_{i-1}/V(\mathbb{D}_i)) = 1$;
(iii) $\mathcal{O}_L$ *acts on* $\mathbb{D}_i$ *via* σ_i;
(iv) *The pairing* $\langle \cdot, \cdot \rangle$ *induces on each* $\mathbb{D}_i$ *a perfect alternating pairing* $\langle \cdot, \cdot \rangle_i$, *and* $\mathbb{D}_i \perp \mathbb{D}_j$ *for* $i \neq j$.

Proof. Part (iii) follows straight from the definition.

For clarity, we denote the action of an element $r \in \mathcal{O}_L$ on $x \in \mathbb{D}$ by $[r]x$. Note that for every two elements r and s in $\mathcal{O}_L \otimes W(k)$, we have

$$\langle [r]x, [s]y \rangle = \langle x, [rs]y \rangle.$$

Choosing r, s to be orthogonal idempotents giving the projections onto $\mathbb{D}_i, \mathbb{D}_j$ respectively, (iv) follows.

Let $x \in \mathbb{D}_i$; then $[r]x = \sigma_i(r)x$. Every element $r \in \mathcal{O}_L$ induces an isogeny $[r] : \mathbb{D} \to \mathbb{D}$ of Dieudonné modules. Hence $[r]Fx = F[r]x = F\sigma_i(r)x = \sigma \circ \sigma_i(r)x = \sigma_{i+1}(r)x$. Parts (i) and(ii) follow. □

Definition 2.3.2. An admissible basis for $\mathbb{D}$, given the polarization and the $\mathcal{O}_L$-structure, is a basis $\{X_1, \cdots, X_g, Y_1, \cdots, Y_g\}$ for $\mathbb{D}$, such that
(i) $\{X_i, Y_i\}$ is a symplectic basis for $\mathbb{D}_i$;
(ii) $Y_i \in V(\mathbb{D}_{i+1})$.

Let

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

be the corresponding display for such a basis. Note that for suitable a_i, b_i, c_i, d_i, we have

$$A = \mathfrak{d}_2(a_2, \cdots, a_g, a_1), \quad B = \mathfrak{d}_2(b_2, \cdots, b_g, b_1),$$
$$C = \mathfrak{d}_2(c_2, \cdots, c_g, c_1), \quad D = \mathfrak{d}_2(d_2, \cdots, d_g, d_1).$$

We have a universal display over $\mathcal{C}(R)$ as in (2.3).

Proposition 2.3.3. *Let* $R = k[[t_{ij}]]/(t_{ij} - t_{ji})$. *Then the maximal closed subscheme of* $\mathrm{Spec}(R)$ *to which the* $\mathcal{O}_L$*-action and the level-n structure extends is given by* $\mathcal{H} = \mathrm{Spec}(H)$, *where* $H = R/(t_{ij} : i \neq j)$.

Proof. Since n is prime to p, the level structure extends to any deformation, and uniquely. We can therefore disregard it.

It is enough to prove that the $\mathcal{O}_L$-action extends to this closed subscheme, because it is well known that the local deformation ring of the Hilbert modular variety is regular on g parameters. Let $\mathbb{D}_H$ be the display derived from (2.3) by putting $t_{ij} = 0$ for $i \neq j$. Then

$$\mathbb{D}_H = \bigoplus \mathbb{D}_{H,i},$$

where $\mathbb{D}_{H,i}$ is obtained from $\mathbb{D}_i$ by extending scalars to $W(H)$, with the naturally given action of $W(k)$ on each component. The action of $\mathcal{O}_L$ is defined via the map

$$\mathcal{O}_L \longrightarrow \bigoplus_i W(k)\ , \quad a \mapsto (\sigma_1(a), \cdots, \sigma_g(a)).$$

This is a map of $\mathcal{C}(H)$-modules if and only if it commutes with the Frobenius; that is, if and only if

$$M_1 M_2^{\sigma} = M_2 M_1, \tag{2.6}$$

where

$$M_1 = \begin{pmatrix} A+TC & p(B+TC) \\ C & pD \end{pmatrix},$$

$$M_2 = \begin{pmatrix} \mathfrak{d}_1(\sigma_1(a), \cdots, \sigma_g(a)) & 0 \\ 0 & \mathfrak{d}_1(\sigma_1(a), \cdots, \sigma_g(a)) \end{pmatrix}.$$

It is easy to verify that (2.6) holds. □

Theorem 2.3.4. *Let $t \in \mathcal{M}_n(k)$ be a geometric point of type τ. There is a choice of parameters $t_1, \cdots, t_g$, such that the universal local deformation ring of $\mathbf{X}_t$ is isomorphic to* Spec$(k[[t_1, \cdots, t_g]])$, *and such that for every type $\rho \subseteq \tau$ the property of "being of type containing ρ" is given by the closed regular $(g - |\rho|)$-dimensional formal subscheme defined by the ideal $(t_i : i \in \rho)$.*

Proof. The universal Dieudonné module $\mathbb{D}_H$ is displayed by the matrix

$$\begin{pmatrix} A+TC & B+TC \\ C & D \end{pmatrix},$$

and

$$A + TC = \mathfrak{d}_2(a_2 + T_2 c_2, \cdots, a_n + T_g c_g, a_1 + T_1 c_1),$$

$$C = \mathfrak{d}_2(c_2, \cdots, c_g, c_1).$$

By Remark 2.2.8, the matrix

$$\begin{pmatrix} A+TC & 0 \\ C & 0 \end{pmatrix} \pmod p$$

is the matrix of the Frobenius on $H_1^{dR}(\mathfrak{X})$, where $\mathfrak{X}$ is the universal deformation over $\mathrm{Spec}(k[[t_1, \cdots, t_g]])$, and the matrix

$$A + TC = \mathfrak{d}_2(a_2 + c_2T_2, \cdots, a_g + c_gT_g, a_1 + c_1T_1),$$

read mod p, is the Hasse-Witt matrix. Note that $a_i = 0 \pmod p$ if and only if $F(\mathbb{D}_{i-1}) = V(\mathbb{D}_{i+1})$, if and only if $i \in \tau(X)$. Since $a_i \neq 0 \pmod p$ implies that $a_i + t_ic_i$ is invertible $(\mathrm{mod}\, p)$, the theorem follows. □

Definition 2.3.5. For $\tau \subseteq \mathbb{Z}/g\mathbb{Z}$, let W_τ be the closed subscheme of $\mathcal{M}_n$, with the reduced induced structure, having the property that for every geometric point $t \in W_\tau$ we have $\tau(\mathbf{X}_t) \supseteq \tau$. We let

$$W_\tau^0 = W_\tau \setminus \bigcup_{\tau' \supset \tau, \tau' \neq \tau} W_{\tau'}.$$

Corollary 2.3.6. *The subscheme W_τ is a locally irreducible and regular scheme. In particular, different components of W_τ do not intersect and every component is regular.*

Proof. This follows directly from the theorem. □

Corollary 2.3.7. *For any two types τ and ρ,*

$$W_\tau \cap W_\rho = W_{\tau \cup \rho}.$$

Proof. Clearly we have equality at least on geometric points, which implies that $W_\tau \cap W_\rho \supseteq W_{\tau\cup\rho}$. The only question is of multiplicity, and this is supplied by Theorem 2.3.4. □

Definition 2.3.8. Let $\mathfrak{X}/\mathcal{M}_n$ be the universal principally polarized abelian scheme with real multiplication by $\mathcal{O}_L$ and symplectic level-n structure. Let $\mathbb{E}$ be the relative cotangent bundle,

$$\mathbb{E} = e^*(\Omega^1_{\mathfrak{X}/\mathcal{M}_n}),$$

where $e : \mathcal{M}_n \to \mathfrak{X}$ is the zero section. Let λ_1 be the first Chern class of $\mathbb{E}$, or, equivalently, of $\det(\mathbb{E})$.

See Section 4.1 for more on this subject.

Proposition 2.3.9 (compare [EO]). *The variety W_τ^0 is quasi-affine.*

Proof. It is well known that λ_1 is an ample line bundle on $\mathcal{M}_n$. Indeed, the sections of $k\lambda_1$ are modular forms of weight k and give the Baily-Borel compactification of $\mathcal{M}_n$, see [vdG], Chapter X, Section 3. Hence λ_1 is ample on W_τ (see also [MB], Introduction and Theorem VII.3.2). We prove next that it is torsion on W_τ^0:

Let $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ be the restriction to W_τ^0 of the universal p-torsion subgroup scheme, i.e., $\mathfrak{X}[p]$, together with its perfect alternating form coming from the polarization on $\mathfrak{X}$. The results of Section 3.2 show that $(\mathcal{G}, \langle\cdot,\cdot\rangle)$ has the same isomorphism type for every geometric point of W_τ^0. It follows that there exists a finite cover $f : T \to W_\tau^0$ such that

$$(\mathcal{G} \times_{W_\tau^0} T, \langle \cdot, \cdot \rangle \times_{W_\tau^0} T)$$

is constant. (One follows the argument of [EO] or [O1], Section 5. See also Remark 3.2.9.) Therefore, its cotangent space,

$$(\mathrm{Lie}(\mathcal{G} \times_{W_\tau^0} T))^* \cong (\mathbb{E}|_{W_\tau^0}) \times_{W_\tau^0} T,$$

is constant as well, and so is its determinant. Hence $f^*\lambda_1$ is trivial, and since f is finite, λ_1 is torsion.

We conclude that for certain $m \geq 1$, the line bundle $m\lambda_1$ is trivial on W_τ^0 and very ample on W_τ. It follows that we can represent $m\lambda_1$ by an effective divisor D contained in $W_\tau - W_\tau^0$. Hence $W_\tau - D$ is quasi-affine, and so is W_τ^0. □

Before stating the next corollary, it may benefit the discussion to formulate the following principle which we use repeatedly below.

Remark 2.3.10. Let S be an integral scheme and let ξ be its generic point. Let t be a nonsingular geometric point of S. Let X/S be an abelian scheme, and denote by X_ξ and X_t the corresponding fibres. We may read many of the properties of X_ξ from the local deformation theory of X_t in the following manner.

Let T be the formal neighborhood of t in S. Let η be the generic point of T; then X_η is obtained from X_ξ by base change of fields. Hence, e.g., X_η and X_ξ have the same a-number, f-number, Newton polygon, etc.

Corollary 2.3.11. *The generic point of every component of W_τ has type τ.*

Proof. Let C be a component of W_τ. It follows from Proposition 2.3.9 that C contains a point c of a strictly larger type τ'. It follows from Theorem 2.3.4 that we can find a local deformation of c whose generic point is of type τ. Now, using the local irreducibility of C, given by Corollary 2.3.6, the assertion follows.

Corollary 2.3.12. *Every component of W_τ contains a superspecial point.*

Proof. Let C be a component of W_τ. It follows from Proposition 2.3.9 that C contains a point c of a strictly larger type τ'. From Theorem 2.3.4 we know that locally at c the subscheme $W_{\tau'}$ is irreducible and is contained in W_τ. It follows that W_τ contains a component of $W_{\tau'}$, and the result follows by induction.

Corollary 2.3.13. *For every type τ,*

$$\dim(W_\tau) = g - |\tau|.$$

Proof. By Corollary 2.3.12 every component of W_τ contains a superspecial point. Using Theorem 2.3.4 and Corollary 2.3.6, the result follows. □

3. The type and the p-torsion

3.1. The Ekedahl-Oort stratification. Consider the coarse moduli space of principally polarized abelian varieties of dimension g in characteristic p, and denote it by $\mathcal{A}$. Given a geometric point t of $\mathcal{A}$, we get a couple

$$(X_t[p], \langle \cdot, \cdot \rangle),$$

where X_t is the abelian variety parameterized by t and

$$\langle \cdot, \cdot \rangle : X_t[p] \times X_t[p] \to \mu_p$$

is the perfect alternating pairing obtained from the polarization λ_t on X_t.

Ekedahl and Oort proved that over an algebraically closed field there are finitely many isomorphism classes of such couples, and that every isomorphism class is locally closed. The stratification itself is formed by taking carefully chosen unions of such locally closed sets. Since the details are a bit lengthy, we refer the reader to [O1], [EO], and [O2] for more on this fascinating story.

3.2. The type determines the Ekedahl-Oort strata in $\mathcal{M}_n$.

Circular groups, swords, and dwords.

We follow [KO] and [O2], extracting what we need.

Let k be an algebraically closed field of characteristic p. Let $w = a_1 \cdots a_s$ be a word in the letters $\{F, V\}$, i.e., a_i is an element of $\{F, V\}$. Define

$$M_w = \bigoplus_{i=1}^{s} k \cdot x_i,$$

and make M_w into a (contravariant) Dieudonné module by defining

$$Fx_i = 0, \;\; x_i = Vx_{i+1} \quad \text{if } a_i = V,$$

$$Fx_i = x_{i+1}, \;\; Vx_{i+1} = 0 \quad \text{if } a_i = F,$$

where by x_{s+1} we mean x_1.

Thinking about the generators as the hours marks on a clock, F acts clockwise and V acts counterclockwise, where $a_i = F$ (resp., $a_i = V$) is telling us that F is not zero (resp., is zero) on x_i. Note that the action of V is then "almost imposed on us", if we wish to have $\mathrm{Im}(V) = \mathrm{Ker}(F)$ and $\mathrm{Im}(F) = \mathrm{Ker}(V)$. Hence, we call such a Dieudonné module a *circular module* and we call the corresponding group C_w a *circular group*. It is killed by p and indeed satisfies

$$\mathrm{Im}(V) = \mathrm{Ker}(F), \quad \mathrm{Im}(F) = \mathrm{Ker}(V).$$

Note that C_w depends only on w mod cyclic permutations. Another easy property is that $C_w^D \cong C_{w^D}$, where C_w^D is the dual group to C_w, and w^D is the word obtained from w by changing every F to V and every V to F.

A *dword* (dual word) of length $2g$ is a couple (w, w^D), where w is a word of length g. Given a dword $z = (w, w^D)$, we let

$$C_z = C_w \oplus C_{w^D}.$$

A *sword* (special word) is a circular word $w = a_1 \cdots a_{2g}$ with the property $a_i \neq a_{i+g}$ for all i, and for some $1 \leq i \leq 2g$ we have $a_i = V$, $a_{i+1} = F$ (here, as below, the indices are read in a cyclic way).

Example 3.2.1. Consider the sword $w = FVVF$. One finds that C_w is a group scheme of order p^4 killed by F^3, by V^3, and by $FV = p$. It is isomorphic, over an algebraically closed field, to the p-torsion of any supersingular, but not superspecial, abelian surface. If we consider $z = (FV, VF)$ instead, then we get the p-torsion of a superspecial abelian surface. The group C_z is killed by F^2, by V^2, and by $FV = p$. In particular, C_z is not isomorphic to C_w.

Remark 3.2.2. Note that a circular group C is the kernel of multiplication by p on a p-divisible group G, i.e., $G[p] = C$. Conversely, for every p-divisible group G, the kernel of multiplication by p is a direct sum of circular groups. In this respect, our conventions are a slight variation on [O2], Section 13.

We also remark that a circular group C_w is of local-local type if and only if both the letters F and V appear in w. For example, the word $w = VV$ gives the Dieudonné module of μ_p^2, while the word $w = FF$ gives the constant group scheme $(\mathbb{Z}/p\mathbb{Z})^2$.

Lemma 3.2.3. *Let z be a dword or sword of length $2g$. Then C_z has a naturally defined $\mathcal{O}_L$-structure and $\mathcal{O}_L$-polarization, i.e., a perfect alternating pairing*

$$\langle \cdot, \cdot \rangle : C_z \times C_z \longrightarrow k$$

such that

$$\langle Fx, y \rangle = \langle x, Vy \rangle^\sigma,$$

and

$$\langle ax, y \rangle = \langle x, ay \rangle$$

for every $a \in \mathcal{O}_L$.

Proof. We indicate how to define the pairing and $\mathcal{O}_L$-structure, leaving the verification of the details to the reader.

Assume that $z = (w, w^D)$ is a dword and that

$$M_w = \bigoplus_{i=1}^{g} k \cdot x_i, \quad M_{w^D} = \bigoplus_{i=1}^{g} k \cdot y_i.$$

We let

$$M_z = \bigoplus_{i=1}^{g} (M_z)_i, \quad (M_z)_i = k \cdot x_i + k \cdot y_i,$$

and note that it has a natural $\mathcal{O}_L \otimes k \cong \bigoplus_i W(k)$-structure and hence, via

$$\mathcal{O}_L \hookrightarrow \mathcal{O}_L \otimes k, \;\; a \mapsto (\sigma_1(a), \cdots, \sigma_g(a)),$$

also an $\mathcal{O}_L$-structure. The pairing is defined by decreeing

$$\langle x_i, y_i \rangle = -\langle y_i, x_i \rangle = 1,$$

and

$$(M_z)_i \perp (M_z)_j, \;\; i \neq j.$$

If z is a sword, we basically do the same. We let

$$M = \bigoplus_{i=1}^{2g} k \cdot x_i,$$

and we put

$$(M_w)_i = k \cdot x_i + k \cdot x_{i+g},$$

$$\langle x_i, x_{i+g} \rangle = 1, \;\; 1 \leq i \leq g,$$

etc. □

Remark 3.2.4. Note that if z is a sword then the $\mathcal{O}_L$-structure on C_z depends on z and not only on z up to cyclic permutations. The same remark applies to dwords.

Proposition 3.2.5. 1. *Let $z = (w, w^D)$ be a dword; then C_w is $\mathcal{O}_L$-invariant and indecomposable as an $\mathcal{O}_L$-group. In this case, the decomposition $C_z = C_w \oplus C_{w^D}$ is the unique decomposition, up to isomorphism, of C_z into indecomposable $\mathcal{O}_L$-groups.*

2. *Let z be a sword. Then C_z is indecomposable as an $\mathcal{O}_L$-group.*

Remark 3.2.6. Note that we are not claiming that C_z or C_w are simple $\mathcal{O}_L$-groups. This is not the case. For example, when g equals 2, a supersingular point t of $\mathcal{M}_n$ with a-number equal to 1 gives a counterexample. In this case $X_t[p]$, which by Theorem 3.2.8 below is C_z for a suitable sword z – in fact z is $FVVF$ – has a family of $\mathcal{O}_L$-subgroups of order p^2 parameterized by $\mathbb{P}^1$.

Proof. 1. Assume that $z = (w, w^D)$ is a dword, where $w = a_1 \cdots a_g$. It is clear that C_w is $\mathcal{O}_L$-invariant. Assume that $C_w = H \oplus K$ is a nontrivial decomposition into $\mathcal{O}_L$-groups. We look at the Dieudonné modules:

$$M_w = \bigoplus_{i=1}^{g} (M_w)_i = \bigoplus_{i=1}^{g} k \cdot x_i = \bigoplus_{i=1}^{g} (\mathcal{H}_i \oplus \mathcal{K}_i),$$

where $\mathcal{H} = \mathcal{D}(H)$ and $\mathcal{K} = \mathcal{D}(K)$. We see that, for a suitable i, $\mathcal{H}_i = k \cdot x_i$ and $\mathcal{H}_{i+1} = 0$. Thus $\mathcal{K}_i = 0$ and $\mathcal{K}_{i+1} = k \cdot x_{i+1}$. If $a_i = F$, we get $\mathcal{H}_{i+1} = F\mathcal{H}_i \neq 0$, a contradiction. If $a_i = V$, we get $\mathcal{K}_i = V\mathcal{K}_{i+1} \neq 0$, a contradiction.

2. Assume that z is a sword, say $z = a_1 \cdots a_{2g}$. Assume that $C_z = H \oplus K$, a sum of two nontrivial $\mathcal{O}_L$-invariant subgroups, and let $\mathcal{H}$ and $\mathcal{K}$ be the corresponding Dieudonné modules. We have

$$M_w = \bigoplus_{i=1}^{g} M_i = \bigoplus_{i=1}^{g} k \cdot x_i + k \cdot x_{i+g} = \bigoplus_{i=1}^{g} (\mathcal{H}_i \oplus \mathcal{K}_i). \tag{3.1}$$

Let Φ be the kernel of the Frobenius on M_w. It is the image of V. Similarly, the image of F is the kernel of V. For every i, the submodule Φ_i is one dimensional and Φ is $\mathcal{O}_L$-invariant. We establish the following points:

- *For every i, we have both $\mathcal{H}_i \neq 0$ and $\mathcal{K}_i \neq 0$.*

Indeed, if $\mathcal{H}_i \neq 0$ then $\mathcal{H}_{i+1} \neq 0$, because either $F\mathcal{H}_i \neq 0$, or $F\mathcal{H}_i = 0$, and then we have $\mathcal{H}_i = VM_{i+1}$. Then, if $\mathcal{H}_{i+1} = 0$, we get $\mathcal{H}_i = V\mathcal{K}_{i+1}$, a contradiction. Since, at least for one i, $\mathcal{H}_i \neq 0$, we get that this holds in fact for every i and similarly for $\mathcal{K}$.

- *For every i, either $\mathcal{H}_i = \Phi_i$ or $\mathcal{K}_i = \Phi_i$.*

Indeed, either $V\mathcal{H}_{i+1} \neq 0$ or $V\mathcal{K}_{i+1} \neq 0$.

- *Let $\tau = \{1 \leq i_1 < \cdots < i_a \leq g\}$ be the indices such that $FM_{i-1} = VM_{i+1}$. Assume that for some j, $\mathcal{H}_{i_j} = \Phi_{i_j}$; then $\mathcal{K}_{i_{j+1}} = \Phi_{i_{j+1}}$. Similarly, if $\mathcal{K}_{i_j} = \Phi_{i_j}$, then $\mathcal{H}_{i_{j+1}} = \Phi_{i_{j+1}}$.*

The idea is this. Since $\mathcal{H}_{i_j} = \Phi_{i_j}$, we must have $F\mathcal{K}_{i_j} \neq 0$, and hence it is the kernel of V on $(M_w)_{i_j+1}$. If $i_{j+1} \neq i_j + 1$, then $F\mathcal{K}_{i_j} \neq VM_{i_j+2} = \Phi_{i_j+1}$.

Thus, $F^2\mathcal{K}_{i_j} \neq 0$, etc. This stops exactly at the first s such that $F^s\mathcal{K} \in VM$. Thus, $\mathcal{K}_{i_{j+1}} = \Phi_{i_{j+1}}$.

To finish the proof of irreducibility we just note that for any sword w, $a(C_w)$ is odd (then the last • gives the contradiction). This is the combinatorics.

We assume, to simplify notation, that in the word $w = a_1 \cdots a_{2g}$ we have $a_1 = F$ and $a_{2g} = V$. Note that $a(C_w)$ is the number of times that VF appears in the circular word w. More precisely, $i+1$ belongs to the type if and only if $a_i \neq a_{i+1}$. Write the elements of w in two rows:

a_1	a_2	a_3	$\cdots$	a_g
a_{g+1}	a_{g+2}	a_{g+3}	$\cdots$	a_{2g}

Go along this brick road from left to right, jumping from a V brick to a V brick. You start at a_{g+1} and end at a_{2g}. The number of times you jump to the upper row is thus the number of times you jump back to the lower row. On the other hand each jump counts a couple VF – either on the upper or on the lower row (but just on one of them each time). Thus we get an even number of couples VF. But we still have to consider $a_g a_{g+1}$ and $a_{2g}a_1$, which are just FV and VF respectively. Hence the total number of couples VF is odd.

Finally, our method of proof shows that in fact the decomposition into indecomposable $\mathcal{O}_L$-groups is unique up to the obvious isomorphisms. □

Lemma 3.2.7. *Let z and t be each either a sword or a dword. Then $C_z \cong C_t$ as $\mathcal{O}_L$-groups if and only if both z and t are swords and $z = t$ or $z = t^D$, or both $z = (w, w^D)$ and $t = (v, v^D)$ and $w = v$ or $w = v^D$.*

Proof. If we have such equalities, clearly we get isomorphic polarized $\mathcal{O}_L$-groups (by construction). Conversely, Proposition 3.2.5 implies that z and t are either both swords or both dwords. If both are dwords, we are done because, obviously, C_w, together with the decomposition into one-dimensional spaces $M_w = \bigoplus M_i$, determines w. We are thus left with the case of swords.

An isomorphism as $\mathcal{O}_L$-groups

$$C_z \cong C_t$$

implies an isomorphism

$$h = \bigoplus h_i : \bigoplus (M_z)_i \to \bigoplus (M_t)_i,$$

where $(M_z)_i, (M_t)_i$ are as in (3.1). Say $w = a_1 \cdots a_{2g}$ and $t = b_1 \cdots b_{2g}$. For some i we are given that $a_i = V, a_{i+1} = F$. This implies that the $i+1$

component of

$$C_z/(FC_z + VC_z) \cong C_t/(FC_t + VC_t)$$

is nonzero ("$i+1$ belongs to the type"). Thus $b_i b_{i+1} = VF$ or $b_{i+g} b_{i+g+1} = VF$.

Assume the first case holds, else take t^D. The idea is that both words are strings of letters – having the same letter in the i-th place – such that the letter changes, say $a_j \neq a_{j+1}$, if and only if either $a_j a_{j+1} = FV$, or $a_j a_{j+1} = VF$, that is, if and only if $a_j a_{j+1} = VF$, or $a_{j+g} a_{j+g+1} = VF$—equivalently, if and only if the $j+1$ component of

$$C_z/(FC_z + VC_z) \cong C_t/(FC_t + VC_t)$$

is nonzero ("$j+1$ belongs to the type"). Hence z equals t. □

The type determines the Ekedahl-Oort stratification in $\mathcal{M}_n$.

We prove that the type determines the Ekedahl-Oort stratification. More precisely,

Theorem 3.2.8. *Let $t \in \mathcal{M}_n(k)$ be a geometric point.*

1. *Suppose that $|\tau(X_t)|$ is odd. There exists a unique sword w, depending only on $|\tau(X_t)|$, such that $X_t[p]$ is isomorphic to C_w with the $\mathcal{O}_L$-structure and polarization.*

2. *Suppose that $|\tau(X_t)|$ is even. There exists a unique dword z, depending only on $|\tau(X_t)|$, such that $X_t[p]$ is isomorphic to C_z with the $\mathcal{O}_L$-structure and polarization.*

Here uniqueness is taken in the sense of Lemma 3.2.7.

Proof. We may assume that $\tau \neq \emptyset$, since otherwise, we can just take the dword $(F \cdots F, V \cdots V)$. We may also assume that $1 \in \tau$, since this just affects, e.g., for a sword, the place i where $a_i a_{i+1} = VF$.

We first define letters $a_1, \cdots, a_{2g}$. Let

$$\tau = \{1 < i_2 < \cdots < i_a \leq g\}.$$

We let

$$(a_1, \cdots, a_{i_2-1}) = (F, \cdots, F), \ \ (a_{i_2}, \cdots, a_{i_3-1}) = (V, \cdots, V), \cdots,$$

$$(a_{i_a}, \cdots, a_g) = (X, \cdots, X),$$

where $X = F$ or V (depending of course on the parity of τ). Then, we just define $a_{i+g} = F$ if $a_i = V$, and $a_{i+g} = V$ if $a_i = F$.

We consider the case of $|\tau(X_t)| = a$ even. In this case X equals V. Let w be the word $a_1 \cdots a_g$ and let z be the dword (w, w^D). Proving an isomorphism $\mathcal{D}(X_t[p]) \cong M_z$ amounts to providing, for every i, a symplectic basis x_i, y_i of $\mathcal{D}(X_t[p])_i$, with the property

$$Fx_i = x_{i+1}, \quad y_i = Vy_{i+1},$$

if $a_i = F$, and

$$x_i = Vx_{i+1}, \quad Fy_i = y_{i+1},$$

if $a_i = V$.

In what follows x_i, y_i, will denote elements in $\mathcal{D}(X_t[p])_i$. That may very well involve a tacit claim that this is possible! We choose the x_i, y_i in the following way.

Start by taking some nonzero $y_1 \in \mathrm{Ker}(F)$ and x_1 such that $\langle x_1, y_1\rangle = 1$. Hence $x_1 \notin \mathrm{Ker}(F)$. Then choose:

definition of x_i	definition of y_i	$a_i a_{i+g}$	i
x_1	$y_1 \in \ker(F)$	FV	1
$x_2 = Fx_1$	$y_2 \in \ker(F) : Vy_2 = y_1$	FV	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{i_2-1} = Fx_{i_2-2}$	$y_{i_2-1} \in \ker(F) : Vy_{i_2-1} = y_{i_2-2}$	FV	$i_2 - 1$
$x_{i_2} = Fx_{i_2-1}$	$y_{i_2} : Vy_{i_2} = y_{i_2-1}$	VF	i_2
$x_{i_2+1} \in \ker(F) : Vx_{i_2+1} = x_{i_2}$	$y_{i_2+1} = Fy_{i_2}$	VF	$i_2 + 1$
$x_{i_2+2} \in \ker(F) : Vx_{i_2+2} = x_{i_2+1}$	$y_{i_2+2} = Fy_{i_2+1}$	VF	$i_2 + 2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
		VF	$i_3 - 1$
$x_{i_3} : Vx_{i_3} = x_{i_3-1}$	$y_{i_3} = Fy_{i_3-1}$	FV	i_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{i_a} = Fx_{i_a-1}$	$y_{i_a} : Vy_{i_a} = y_{i_a-1}$	VF	i_a
$x_{i_a+1} \in \ker(F) : Vx_{i_a+1} = x_{i_a}$	$y_{i_a+1} = Fy_{i_a}$	VF	i_{a+1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_g \in \ker(F) : Vx_g = x_{g-1}$	$y_g = Fy_{g-1}$	VF	g
$x_1' : Vx_1' = x_g$	$y_1' = Fy_g$		

Note that for every r we have, by induction, $\langle x_r, y_r\rangle = 1$. Hence, x_r and y_r form a symplectic basis to $\mathcal{D}(X_t[p])_r$. Note that

$$y_1' \in \mathrm{Im}\left(F|_{\mathcal{D}(X_t[p])_g}\right) = \mathrm{Ker}\left(V|_{\mathcal{D}(X_t[p])_1}\right) = \mathrm{Ker}\left(F|_{\mathcal{D}(X_t[p])_1}\right).$$

Therefore, since k is algebraically closed, by taking a suitable multiple of y_1 we may assume that $y_1' = y_1$.

Since $\langle x_1', y_1 \rangle = \langle x_1, y_1 \rangle = 1$ we have $x_1 - x_1' \in \mathrm{Ker}(F)$. Thus we may also assume that $x_1' = x_1$. This finishes the proof in the case of even a. The case of odd a is similar enough to be left to the reader. We just hint that the required word w is $a_1 \cdots a_{2g}$, the construction goes the same, and one argues that by a suitable modification we can get $x_1' = y_1$, $y_1' = x_1$. □

Remark 3.2.9. The proof of Theorem 3.2.8 suggests another method of proving the assertion, made in the proof of Proposition 2.3.9, about the existence of a finite covering $T \longrightarrow W_\tau^0$ such that, denoting by $\mathcal{G}$ the universal p-torsion subgroup scheme over W_τ^0,

$$\mathcal{G} \underset{W_\tau^0}{\times} T$$

is constant. The problem, in terms of the proof, is to find sections $\{x_i, y_i\}$ satisfying the above relations. Locally on the base, this amounts to taking certain $p^\ell - 1$ roots, for various ℓ's, of functions.

4. The global structure

4.1. The Hodge bundle and W_r. In this section we follow some ideas of [Mu] and [vdG1].

Let

$$\pi : (\mathfrak{X}, \Lambda, I, A) \longrightarrow \mathcal{M}_n$$

be the universal object over $\mathcal{M}_n$. We will abuse notation and write $\mathfrak{X}$ for $(\mathfrak{X}, \Lambda, I, A)$. Let $e : \mathcal{M}_n \to \mathfrak{X}$ be the identity section, and let

$$\mathbb{E} = e^*\left(\Omega^1_{\mathfrak{X}/\mathcal{M}_n}\right) = \pi_*\left(\Omega^1_{\mathfrak{X}/\mathcal{M}_n}\right)$$

be the Hodge bundle over $\mathcal{M}_n$. It is a locally free sheaf of rank g over $\mathcal{M}_n$. Recall that a modular form of (equi-)level k is a section of $\omega^{\otimes k}$, where

$$\omega \overset{\mathrm{def}}{=} \det(\mathbb{E}).$$

It is known that ω is an ample line bundle on $\mathcal{M}_n$—see [vdG], Chapter X, Section 3, and [MB], Introduction, Theorem 1.1.

Consider the subvariety of $\mathcal{M}_n$ defined by

$$W_r \overset{\mathrm{def}}{=} \bigcup_{\tau:\, |\tau| = r} W_\tau.$$

We want to explain how it is related to $c_r(\mathbb{E})$, where c_r denotes the r-th Chern class of a vector bundle. This gives a "global interpretation" to our stratification. To make our relations exact we continue to take a symplectic full level-n structure, where n is at least 3 and prime to p, and work with $\mathcal{M}_n$. The result for the coarse moduli space $\mathcal{M}$ – where we take no level structure – involves studying the ramification of the natural morphism from $\mathcal{M}_n$ to $\mathcal{M}$.

To see the problem, notice that when g equals 1, W_1 is the set of supersingular elliptic curves, and $\deg(c_1) = 1/24$. This can be deduced from the following facts:

$$\omega = \mathbb{E}, \quad \omega^{\otimes 2} \cong \Omega^1_{\mathcal{M}_n}(\text{cusps}).$$

The right formula is Deuring's mass formula,

$$\sum_{E \text{ supersingular}} \frac{1}{\mathrm{Aut}(E)} = \frac{p-1}{24} = (p-1)\deg(c_1).$$

Recall that we consider the regular scheme $\mathcal{M}_n$ as a scheme over the finite field $\mathbb{F}$ containing $\mathbb{F}_{p^g}$. In particular we have a decomposition of $\mathcal{O}_L \otimes \underline{\mathcal{O}}_{\mathcal{M}_n}$ induced from the decomposition

$$\mathcal{O}_L/p \otimes \mathbb{F} \cong \bigoplus_{i=1}^{g} \mathbb{F}, \quad a \otimes 1 \mapsto (\sigma_1(a), \cdots, \sigma_g(a)).$$

Thus, over $\mathcal{M}_n$, the Hodge bundle $\mathbb{E}$ is a direct sum of g line bundles:

$$\mathbb{E} = \mathbb{L}_1 \oplus \cdots \oplus \mathbb{L}_g.$$

Given any locally free sheaf $\mathbb{K}$ over $\mathcal{O}_L \otimes \underline{\mathcal{O}}_{\mathcal{M}_n}$, we will denote by $\mathbb{K}_i$ its i-th component. In particular, $\mathbb{L}_i = \mathbb{E}_i$.

Let $\mathfrak{X}^{(p)}$ be the Frobenius image of $\mathfrak{X}$. Recall that it is defined by the following cartesian square:

$$\begin{array}{ccc} \mathfrak{X}^{(p)} & \longrightarrow & \mathfrak{X} \\ \downarrow & & \downarrow \\ \mathcal{M}_n & \xrightarrow{F_{\text{abs}}} & \mathcal{M}_n, \end{array}$$

where F_{abs} is the absolute Frobenius, defined as the identity on the underlying topological space of $\mathcal{M}_n$, and as raising to the p-th power on the structure sheaf.

Let $\mathbb{T}$ be the relative cotangent sheaf of $\mathfrak{X}^{(p)}$ restricted to the origin. It is easy to verify that

$$\mathbb{T} = \mathbb{E}^{(p)},$$

where for any $\underline{\mathcal{O}}_{\mathcal{M}_n}$-sheaf $\mathbb{K}$ we define $\mathbb{K}^{(p)}$ as the $\underline{\mathcal{O}}_{\mathcal{M}_n}$-sheaf obtained by twisting by the Frobenius $\underline{\mathcal{O}}_{\mathcal{M}_n} \to \underline{\mathcal{O}}_{\mathcal{M}_n}$,

$$\mathbb{K}^{(p)} = \underline{\mathcal{O}}_{\mathcal{M}_n} \otimes_{\underline{\mathcal{O}}_{\mathcal{M}_n}} \mathbb{K}.$$

The definition is compatible under the dictionary between locally free sheaves and vector bundles.

The Verschiebung morphism

$$\mathfrak{X}^{(p)}/\mathcal{M}_n \longrightarrow \mathfrak{X}/\mathcal{M}_n,$$

which is $\mathcal{O}_L$-equivariant, induces a *linear* $(\mathcal{O}_L \otimes \underline{\mathcal{O}}_{\mathcal{M}_n})$-map

$$V : \mathbb{E} \longrightarrow \mathbb{T}.$$

This can be justified either by interpreting $\mathbb{E}$ and $\mathbb{T}$ as cohomology objects via their embedding in the first de Rham cohomology, or by interpreting them as the cotangent sheaves. We thus get, for every i, an $(\mathcal{O}_L \otimes \underline{\mathcal{O}}_{\mathcal{M}_n})$-linear map

$$V : \mathbb{E}_i \longrightarrow \mathbb{T}_i,$$

where $\mathcal{O}_L$ acts via σ_i.

Since $\mathbb{E} = \bigoplus \mathbb{E}_i$, it follows from the definitions that

$$\mathbb{T}_i = \mathbb{E}_{i-1}^{(p)}.$$

Thus, we get an $(\mathcal{O}_L \otimes \underline{\mathcal{O}}_{\mathcal{M}_n})$-linear map

$$V : \mathbb{E}_i \longrightarrow \mathbb{E}_{i-1}^{(p)}.$$

Furthermore, over any perfect $\underline{\mathcal{O}}_{\mathcal{M}_n}$-algebra R, in particular, for every geometric point of $\mathcal{M}_n$, we get a σ^{-1}-linear map,

$$V : \mathbb{E}_i \longrightarrow \mathbb{E}_{i-1},$$

by twisting with respect to σ^{-1}. That is, over R,

$$\mathbb{E}_{i-1} = \mathbb{E}_{i-1}^{(p)} \otimes_R R.$$

Therefore, from this point of view, $W_{\{i-1\}}$ is the degeneracy locus of the linear map

$$V : \mathbb{E}_i \longrightarrow \mathbb{E}_{i-1}^{(p)}.$$

To wit, for every geometric point t,

$$(\mathcal{D}(\alpha(\mathfrak{X}|_t)))_i = \left(H^0\left(\Omega^1_{\mathfrak{X}|_t}\right)/VH^0\left(\Omega^1_{\mathfrak{X}|_t}\right)\right)_i = (\mathbb{E}_i|_{\mathfrak{X}_t})\,/\,(V\mathbb{E}_{i+1}|_{\mathfrak{X}_t}).$$

Thus, i belongs to $\tau(\mathfrak{X}_t)$ if and only if

$$V : \mathbb{E}_{i+1}|_{\mathfrak{X}_t} \longrightarrow \mathbb{E}_i|_{\mathfrak{X}_t}$$

is the zero map.

Let

$$\ell_i = c_1(\mathbb{L}_i)$$

be the first Chern class of the line bundle $\mathbb{L}_i$; then

$$c_r(\mathbb{E}) = \sum_{1\le i_1<\cdots<i_r\le g} \ell_{i_1}\cdots\ell_{i_r}. \tag{4.1}$$

Applying the Thom-Porteous formula, see [F], Chapter 14, we conclude that in the Chow ring $CH^*_{\mathbb{Q}}(\mathcal{M}_n)$,

$$\begin{aligned} W_{\{i\}} &= c_1(\mathbb{L}_i^{(p)} - \mathbb{L}_{i+1}) \\ &= c_1(\mathbb{L}_i^{\otimes p} - \mathbb{L}_{i+1}) \\ &= pc_1(\mathbb{L}_i) - c_1(\mathbb{L}_{i+1}) \\ &= p\ell_i - \ell_{i+1}. \end{aligned}$$

Using Corollary 2.3.7, we conclude the following theorem.

Theorem 4.1.1. *The following equalities hold in* $CH^*_{\mathbb{Q}}(\mathcal{M}_n)$*:*

$$W_{\{i_1,\cdots,i_r\}} = \prod_{j=1}^{r}\left(p\ell_{i_j} - \ell_{i_j+1}\right),$$

$$W_r = \sum_{\substack{I\subset\mathbb{Z}/g\mathbb{Z}\\ |I|=r}} \prod_{j\in I}(p\ell_j - \ell_{j+1}).$$

Corollary 4.1.2. *In* $CH^*_{\mathbb{Q}}(\mathcal{M}_n)$,

$$W_1 = (p-1)c_1(\mathbb{E}).$$

Corollary 4.1.3. *The subspace of* $CH^*_{\mathbb{Q}}(\mathcal{M}_n)$ *spanned by the* tautological classes $\{\ell_1,\cdots,\ell_g\}$ *is equal to the one spanned by* $\{W_1,\cdots,W_g\}$.

4.2. A simplicial complex. In this section, we show that the components of the W_τ's form a simplicial complex, where every top-dimensional simplex is of dimension $g-1$ and corresponds to a superspecial point.

Definition 4.2.1. Define a graph $\mathcal{T}$ with $g+1$ levels, whose vertices at the i-th level are the irreducible components of the W_τ's with $|\tau| = i$. Edges in this graph exist only between consecutive levels. Two vertices v, v', of levels i and $i+1$ respectively, corresponding to components C and C' are connected if and only if $C \supset C'$.

Theorem 4.2.2. *The graph* $\mathcal{T}$ *is a colored simplicial complex, in which the faces of dimension* i *are the vertices of level* $i+1$ *in* $\mathcal{T}$. *In particular,* $\mathcal{M}_n$*, which is equal to* $W_\emptyset$*, corresponds to the empty set and the superspecial locus of* $\mathcal{M}_n$ *corresponds to the* $(g-1)$*-dimensional simplices. Every maximal* *simplex is of dimension $g-1$. The vertices are colored according to their type, and simplices intersect only along faces having the same coloring.*

Remark 4.2.3. One may call $\mathcal{T}$ the "intersection complex" of all the W_τ's.

Proof. All the assertions follow from the following points:

• *The intersection of a component of W_τ and a component of W_σ is either empty or a component of $W_{\tau\cup\sigma}$.*

This follows from Corollary 2.3.7.

• *For every τ, every component C of W_τ and every $\tau' \subsetneq \tau$ there exists a unique component C' of $W_{\tau'}$ such that $C' \supseteq C$.*

Take any geometric point t of C. Theorem 2.3.4 shows that we may deform $\mathbf{X}_t = (X_t, \lambda_t, \iota_t, \alpha_t)$ to a quadruple $\mathbf{X} = (X, \lambda, \iota, \alpha)$ such that $\tau(\mathbf{X}) = \tau'$. Hence t lies on some component C' of $W_{\tau'}$. But the local irreducibility of $W_{\tau'}$, given by Corollary 2.3.6, implies C' is unique.

• *For every τ, every component C of W_τ and every $\tau' \supseteq \tau$, there is a component C' of $W_{\tau'}$ such that $C' \subseteq C$.*

By Corollary 2.3.12, C has a superspecial point. Use the previous point, or repeat the argument. □

Corollary 4.2.4. *Let H be a family of prime-to-p Hecke operators acting transitively on the superspecial points in $\mathcal{M}_n$. Then H acts transitively on the components of W_τ for every τ.*

Proof. The prime-to-p Hecke operators act on the colored simplicial complex $\mathcal{T}$, i.e., they also preserve the type. The result follows by descending induction on the dimension.

By assumption it holds for the top simplices. Assume it holds for faces of dimension $i+1$ and take two faces of dimension i with the same color, i.e., two components C_1 and C_2 of W_τ, where $|\tau|$ equals i. Let

$$\ell \notin \tau, \quad \tau' = \tau \cup \ell.$$

Let D_1 and D_2 be components of type τ', such that

$$D_j \subseteq C_j, \quad j = 1, 2.$$

By induction, there exists a Hecke operator T such that $D_1 \in T(D_2)$; hence $T(C_2)$ contains a component C_1' of W_τ containing D_1. Since there is a unique such component, we get that $C_1' = C_1$. □

It is thus natural to study the representation of the prime-to-p Hecke algebra on the complex $\mathcal{T}$ as well as its homology as an abstract simplicial complex! This could be viewed as a generalization of the classical study of the action of the Hecke operators on the zero-dimensional complex of supersingular elliptic curves (the case $g = 1$ in our discussion).

As an indication of the interesting information that this complex encrypts, we mention the following.

Theorem 4.2.5 (See [BG1]). *Let $g = 2$. Assume that $p > 2$ and $n \geq 3$. The number of components of $W_{\{i\}}$, for i equal to 1 or 2, is*

$$\tfrac{1}{2}[\mathcal{M}_n : \mathcal{M}] \cdot \zeta_L(-1),$$

and the number of components of $W_{\{1,2\}}$ is

$$\frac{p^2+1}{2}[\mathcal{M}_n : \mathcal{M}] \cdot \zeta_L(-1).$$

5. Newton polygons

5.1. This section is devoted to some results concerning the Newton polygon stratification of $\mathcal{M}_n$ and a conjecture on the dimensions of the various strata. We hope to return to this conjecture in a subsequent paper. Below, we give certain partial results, which imply, in particular, what one may call the "weak Grothendieck conjecture for $\mathcal{M}_n$":

Let β and γ be two symmetric Newton polygons with the same initial and end points $(0,0)$, $(2g, g)$, respectively. Assume that $\beta \leq \gamma$ (see Definition 5.2.3 below). The weak Grothendieck conjecture, for abelian varieties, asserts that there exist an integral scheme S, with a generic point η and a geometric point $t \in S$, and an abelian scheme $\mathfrak{X}$ over S, such that the Newton polygon of $\mathfrak{X}_\eta$ is β and the Newton polygon of $\mathfrak{X}_s$ is γ. The strong Grothendieck conjecture is obtained by specifying the isomorphism class of $\mathfrak{X}_t$.

It is clear how to formulate such conjectures for any PEL Shimura variety (though not their validity), under the obvious assumption that β and γ appear for some geometric points of this Shimura variety, or how to formulate it for p-divisible groups. Theorem 5.3.3 below implies that the weak Grothendieck conjecture holds for the Hilbert modular variety $\mathcal{M}_n$.

We remark that Grothendieck formulated his conjecture, in the setting of p-divisible groups, in a letter to Barsotti, reproduced as the Appendix to [Gr]. Given Theorem 2.2.3, this is essentially the "weak Grothendieck conjecture" for abelian varieties. A stronger conjecture, regarding a sequence of polygons, was formulated by Koblitz in [K], p. 215. For more extensive discussion and results in this direction, including proofs of the "weak versions", we refer the reader to [O3], Section 6.

Clearly, the Grothendieck conjecture is about the variation of the Newton polygon along the local deformation ring of an abelian variety (possibly with some extra structure). It is clear therefore that any local approach to this conjecture is intimately related to Conjecture 5.2.5 (below).

5.2. The possible p-divisible groups.

Theorem 5.2.1. *Let t be a geometric point of $\mathcal{M}_n$. The p-divisible group $X_t(p)$ is isogenous to*

$$G_{\ell/g} + G_{(g-\ell)/g}, \quad 0 \leq \ell \leq g,$$

or

$$g \cdot G_{1/2}$$

(under the convention $ka/kb = k \times a/b$).

Proof. Let t be a geometric point of $\mathcal{M}_n$ and let $(X_t, \lambda_t, \iota_t, \alpha_t)$ be the corresponding object. Let $\mathcal{D}$ be the contravariant Dieudonné module of $X_t(p)$. We have a decomposition as in (2.5):

$$\mathcal{D} = \bigoplus_{i \in \mathbb{Z}/g\mathbb{Z}} \mathcal{D}_i.$$

Note that $(\mathcal{D}, F^g)$ is a σ^g-crystal, which decomposes as a direct sum of g isogenous rank-2 σ^g-crystals,

$$(\mathcal{D}, F^g) = \bigoplus_{i \in \mathbb{Z}/g\mathbb{Z}} (\mathcal{D}_i, F^g).$$

The isogeny is, essentially, F itself. It follows that if the slopes of, say, $(\mathcal{D}_1, F^g)$ are α and β, then the slopes of $(\mathcal{D}, F)$ are α/g and β/g, each appearing with multiplicity g. On the other hand,

$$\mathrm{ord}_p(\bigwedge^2 F^g|_{\mathcal{D}_1}) = p^g,$$

and its slopes have denominator at most 2 (see [Ka], Formula 1.3.1). □

Remark 5.2.2. The converse also holds, i.e., every such isogeny class is obtained from some point of $\mathcal{M}_n$. This follows from Theorem 5.4.11 below.

Definition 5.2.3. We say, for two Newton polygons β, γ, with the same initial and end points, that $\beta \geq \gamma$ if every point of β is not below γ. We say that $\beta > \gamma$ if in addition $\beta \neq \gamma$.

Let $\beta_0 < \beta_1 < \cdots < \beta_s$ be the possible Newton polygons of points of $\mathcal{M}_n$ given by Theorem 5.2.1 (e.g., β_0 is the ordinary polygon and β_s is the supersingular polygon, $s = [(g+1)/2]$). Given a geometric point t of $\mathcal{M}_n$, we let $\mathcal{N}_t$ be the Newton polygon of $(X_t, \lambda_t, \iota_t, \alpha_t)$. For every i such that $0 \leq i \leq s$, let $\mathcal{N}_i$ be the closed subscheme "where $\mathcal{N}_t \geq \beta_i$".

Remark 5.2.4. The reader should note that our definition of the order between Newton polygons is the opposite of a frequently used definition (see for example [O4], Section 3). That is, in our definition $\beta \leq \gamma$ is what is often denoted $\gamma \prec \beta$. Our choice is based on the intuition that under specialization Newton polygons go up and "up" is bigger then "down". The other choice is based on the intuition that if $\gamma \prec \beta$ then the dimension of the Newton strata of γ is less than or equal to that of β.

Conjecture 5.2.5. $\dim(\mathcal{N}_i) = g - i$.

Remark 5.2.6. We intend to discuss this conjecture in a future paper. Some partial results of independent interest are given below. See, in particular, Theorems 5.3.3 and 5.4.11.

5.3. The cyclic case. The following lemma is useful to us (see also [O3], Lemma 3.5).

Lemma 5.3.1. *Let k be a perfect field of characteristic p. Let (M, F) be a rank two σ^i-crystal with $\mathrm{ord}_p(\bigwedge^2 F) = g$. Suppose that F is given with respect to some basis by the matrix*

$$m = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}.$$

Then the first Newton slope of (M, F) is given by

(5.1)

$$\min\left\{\frac{g}{2}, \mathrm{ord}_p\left(m_1^{\sigma^i} + m_4\frac{m_3^{\sigma^i}}{m_3}\right)\right\} = \min\left\{\frac{g}{2}, \mathrm{ord}_p\left(m_1 + m_4^{\sigma^i}\frac{m_2}{m_2^{\sigma^i}}\right)\right\}$$

(under the convention $0/0 = 1$).

Proof. Since the argument is symmetric, we just prove the first formula – the one involving m_3.

If $m_3 = 0$, then the result is clear. If $m_3 \neq 0$, then, denoting the basis vectors by X and Y, we may consider the cyclic submodule of finite index generated by X and $Y' = m_1 X + m_3 Y$. With respect to this basis, the Frobenius is given by

$$\begin{pmatrix} 0 & -\frac{m_3^{\sigma^i}}{m_3}(m_1 m_4 - m_2 m_3) \\ 1 & m_1^{\sigma^i} + m_4\frac{m_3^{\sigma^i}}{m_3} \end{pmatrix}.$$

The characteristic polynomial is

$$x^2 - (m_1^{\sigma^i} + m_4\frac{m_3^{\sigma^i}}{m_3})x + \frac{m_3^{\sigma^i}}{m_3}(m_1 m_4 - m_2 m_3).$$

The lemma follows from [Dm], Lemma 2, p. 82, Lemma 3, p. 84 (or rather their proofs). $\square$

Remark 5.3.2. A typical application of Lemma 5.3.1 will be as follows: We will be interested in knowing the variation of the Newton polygon along a certain formal subvariety of the deformation space of $(X, \lambda, \iota, \alpha)$ or of an RM crystal (defined in Section 5.4 below). Using displays, specifically (2.2) and Proposition 2.3.3, we would know the effect of the Frobenius F on a deformation parameterized by this subvariety. Using the argument in the proof of Theorem 5.2.1, it will be enough to study the effect of F^g on some $\mathbb{D}_i$. Here Lemma 5.3.1 becomes handy.

Theorem 5.3.3. *Let $t \in \mathcal{M}_n(k)$ be a geometric point such that $a(X_t) = 1$. Assume that the Newton polygon of X_t is β_ℓ. There is a choice of parameters $t_1, \cdots, t_g$ such that the local equi-characteristic universal deformation ring of X_t is given by $k[[t_1, \cdots, t_g]]$, and such that the following holds:*

For every $r \le \ell$ the closed formal scheme, "where the Newton polygon is weakly above β_r", is given by $t_1 = \cdots = t_r = 0$.

Proof. Let

$$\mathbb{D} = \bigoplus_{i=1}^{g} \mathbb{D}_i$$

be the covariant Dieudonné module of $X_t(p)$. The assumption $a(X_t) = 1$ is equivalent to $\mathbb{D}$ being a cyclic $\mathcal{C}(k)$-module. It is also equivalent to

$$\dim_k(\mathbb{D}/(F\mathbb{D} + V\mathbb{D})) = 1.$$

We assume, without loss of generality, that $\mathbb{D}_1 \not\subset F\mathbb{D}+V\mathbb{D}$. Thus, there exists $x \in \mathbb{D}_1$ that generates $\mathbb{D}$ over $\mathcal{C}(k)$. Consider the sets of elements

$$\{x, V^g x, F^g x\}, \{Fx, V^{g-1}x\}, \cdots, \{F^{g-1}x, Vx\}.$$

Clearly, each generates the respective $\mathbb{D}_i$ over $W(k)[F^g, V^g]$. In fact, for every $i \neq 1$, we have that $\{F^i x, V^{g-i}x\}$ is a basis for $\mathbb{D}_{i+1}$ over $W(k)$, and for $\mathbb{D}_1$ the same holds for $\{X, V^g x\}$. Indeed, to see that, we first reduce to the case of $\mathbb{D}_1$. We note that $\{F^i x, V^{g-i}x\}$ is not a basis if and only if

$$\langle F^i x, V^{g-i}x\rangle \equiv 0 \pmod p.$$

But

$$\langle F^i x, V^{g-i}x\rangle = \langle x, V^g x\rangle^{\sigma^i}.$$

For $\mathbb{D}_1$ we argue as follows. If

$$\langle x, V^g x\rangle \equiv 0 \pmod p,$$

then

$$\langle x, F^g x\rangle = \langle V^g x, x\rangle^{\sigma^g} \equiv 0 \pmod p,$$

and hence x, $F^g x$, and $V^g x$ do not span $\mathbb{D}_1$ over $W(k)$. This is a contradiction, because they span over $W(k)[F^g, V^g]$ and note what F^g, V^g do to the spanning set.

It follows that for k algebraically closed, we can normalize x such that

$$\langle x, V^g x\rangle = 1.$$

Hence,

$$X_1 = x, Y_1 = V^g x, X_2 = Fx, Y_2 = V^{g-1}x, \cdots, X_g = F^{g-1}X, Y_g = Vx,$$

is an admissible basis for $\mathbb{D}$ as defined in 2.3.2. Note that $F^g X_1 = aX_1 + cY_1$ and

$$c = \langle X_1, F^g X_1\rangle = \langle V^g X_1, X_1\rangle^{\sigma^g} = -1.$$

It follows that, with respect to the basis $\{X_1, \cdots, X_g, Y_1, \cdots, Y_g\}$, the Frobenius is given by

$$\begin{pmatrix} \mathfrak{d}_2(1,1,\cdots,1,a) & \mathfrak{d}_2(0,0,\cdots,0,p) \\ \mathfrak{d}_2(0,0,\cdots,0,-1) & \mathfrak{d}_2(p,p,\cdots,p,0) \end{pmatrix}.$$

The universal deformation over $\mathrm{Spec}(k[[t_1,\cdots,t_g]])$ is given, with respect to the basis $\{X_1, Y_1, \cdots, X_g, Y_g\}$, by

$$\mathfrak{d}_2(B_2, \cdots, B_g, B_1),$$

where

$$B_2 = \begin{pmatrix} 1 & pT_2 \\ 0 & p \end{pmatrix}, \cdots, B_g = \begin{pmatrix} 1 & pT_g \\ 0 & p \end{pmatrix}, B_1 = \begin{pmatrix} a - T_1 & p \\ -1 & 0 \end{pmatrix}.$$

Remark 5.3.4. If g_i are $\sigma^{n(i)}$-linear maps, given, respectively, by matrices $M(i)$, then $g_1 \circ g_2$ is a $\sigma^{n(1)+n(2)}$-linear map and is given by the matrix $M_1 M_2^{\sigma^{n(1)}}$.

We compute the deformation of $(\mathbb{D}_1, F^g)$ resulting from the universal deformation. Put $S_1 = T_1$ and else $S_i = T_i^{\sigma^{g-i+1}}$:

$$\begin{pmatrix} a - S_1 & p \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & pS_g \\ 0 & p \end{pmatrix}\cdots\begin{pmatrix} 1 & pS_2 \\ 0 & p \end{pmatrix}$$
$$= \begin{pmatrix} a - S_1 & (a - S_1)(pS_2 + \cdots + p^{g-1}S_g) + p^g \\ -1 & -(pS_2 + \cdots + p^{g-1}S_g) \end{pmatrix}.$$

Clearly, the deformation is ordinary if and only if $S_1 \neq 0$. Assume that $S_1 = 0$; then, using Lemma 5.3.1, we see that the first Newton slope is $\leq \ell = \mathrm{ord}_p(a)$, and for every $k \leq \ell$, the first Newton slope is k if and only if $S_1 = \cdots = S_k = 0$.

Finally, note that S_i is equal to $T_i^{\sigma^{r(i)}}$, where $r(1) = 0$ and, for $2 \leq i \leq g$, $r(i) = g - i + 1$ is the Teichmüller lift of $t_i^{p^{r(i)}}$. □

Corollary 5.3.5. *The strong Grothendieck conjecture for $\mathcal{M}_n$ holds in the case where the special fibre has a-number equal to 1.* □

5.4. The Newton polygon on W_τ. In this section we compute the Newton polygon of the generic fibre of any component of W_τ. We will use Remark 2.3.10; hence we start with an analysis of the superspecial points.

Superspecial crystals.

By a ppRM crystal (for a given L) we mean a σ-crystal $\mathbb{D}$ over $W(k)$ of rank $2g$, where k is an algebraically closed field of characteristic p, together with an action of $\mathcal{O}_L$ and a principal $\mathcal{O}_L$-linear quasi-polarization. We assume that $\mathbb{D}/F\mathbb{D} \cong k^g$; hence we have a σ^{-1}-linear operator V satisfying $VF = FV = p$. Thus, such a $\mathbb{D}$ has a decomposition as in (2.5), and a notion of an admissible basis as in Definition 2.3.2. Note that every principally polarized abelian variety with $\mathcal{O}_L$ action, (X, λ, ι), gives a ppRM crystal $(\mathcal{D}(X(p)), F)$.

We consider a superspecial ppRM crystal $(\mathbb{D}, F)$ over $W(k)$, where k is an algebraically closed field in characteristic p. Spelling this out, it just means that we have

$$F\mathbb{D} = V\mathbb{D},$$

and the Newton slopes of $(\mathbb{D}_1, F^g)$ are $g/2,\ g/2$. Note that if (X, λ, ι) is superspecial, then $(\mathcal{D}(X(p)), F)$ is a superspecial ppRM crystal. Thus, using Corollary 2.3.12, we see that such crystals exist.

Let $\mathbb{D}$ be a superspecial ppRM crystal. Choose any $Y_1 \in V\mathbb{D}_2$ such that there exists $X_1 \in \mathbb{D}_1$ satisfying

$$\langle X_1, Y_1 \rangle = 1,$$

that is, $Y_1 \notin p\mathbb{D}$. We note that any other choice X_1', Y_1' is given by a matrix

$$M = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}. \tag{5.2}$$

To make our notation clear, this means that

$$X_1' = m_1 X_1 + m_3 Y_1, \quad Y_1' = m_2 X_1 + m_4 Y_1.$$

The matrix M satisfies

$$M \in SL_2(W(k)), \qquad M \equiv \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \pmod p. \tag{5.3}$$

Let

$$\begin{aligned} X_2 &= -\frac{1}{p}FY_1,\ Y_2 = FX_1, \\ X_3 &= -\frac{1}{p}FY_2,\ Y_3 = FX_2, \\ &\vdots \\ X_g &= -\frac{1}{p}FY_{g-1},\ Y_g = FX_{g-1}. \end{aligned}$$

Then, inductively, we get

$$\langle X_i, Y_i\rangle = \langle -\frac{1}{p}FY_{i-1}, FX_{i-1}\rangle = -\frac{1}{p}\langle Y_{i-1}, VFX_{i-1}\rangle^\sigma = 1.$$

It follows that $\{X_1, \cdots, X_g, Y_1, \cdots, Y_g\}$ is an admissible basis for $\mathbb{D}$. Let us write

$$FX_g = aX_1 + cY_1, \quad FY_g = bX_1 + dY_1.$$

Then the Frobenius, with respect to the basis $\{X_1, \cdots, X_g, Y_1, \cdots, Y_g\}$, is given by

$$\begin{pmatrix} A & pB \\ C & pD \end{pmatrix} = \begin{pmatrix} \mathfrak{d}_2(0, \cdots, 0, a) & \mathfrak{d}_2(-p, \cdots, -p, b) \\ \mathfrak{d}_2(1, \cdots, 1, c) & \mathfrak{d}_2(0, \cdots, 0, d) \end{pmatrix}.$$

The universal deformation of $\mathbb{D}$ as a polarized ppRM crystal is given by

$$\begin{pmatrix} A + TC & B + TD \\ C & D \end{pmatrix},$$

where $T = \mathfrak{d}_1(T_1, \cdots, T_g)$—that is, by

$$\begin{pmatrix} \mathfrak{d}_2(T_2, T_3, \cdots, T_g, a + cT_1) & \mathfrak{d}_2(-1, -1, \cdots, -1, b' + d'T_1) \\ \mathfrak{d}_2(1, 1, \cdots, 1, c) & \mathfrak{d}_2(0, 0, \cdots, 0, d') \end{pmatrix},$$

where $b = pb'$ and $d = pd'$.

One calculates the effect of F^g of the universal deformation as a ppRM crystal, on $\mathbb{D}_1$. It is

$$\begin{pmatrix} a+cS_1 & b+dS_1 \\ c & d \end{pmatrix} \begin{pmatrix} S_g & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} S_{g-1} & -p \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} S_2 & -p \\ 1 & 0 \end{pmatrix},$$

where $S_1 = T_1, S_g = T_g^\sigma$, etc.

Lemma 5.4.1. *The original F^g on $\mathbb{D}_1$ is given by*

$$Q = \begin{cases} (-1)^{(g-1)/2} p^{(g-1)/2} \begin{pmatrix} a & b \\ c & d \end{pmatrix}, & g \equiv 1 \pmod 2, \\ (-1)^{(g-2)/2} p^{(g-2)/2} \begin{pmatrix} b & -pa \\ d & -pc \end{pmatrix}, & g \equiv 0 \pmod 2. \end{cases}$$

We have:

1. *for g odd: $ad - bc = p$, $p|a, b, d$;*
2. *for g even: $ad - bc = p$, $p|b, d$.*

Conversely, any matrix satisfying these conditions will give a supersingular crystal.

Proof. Consider the case where g is odd. We know that $ad - bc = p$, because F^g is symplectic and of rank p^g on $\mathbb{D}_1$, and we know that $p|b, d$, because $Y_g \in V\mathbb{D}$; hence its image under Frobenius is divisible by p. In order for Q to define a supersingular crystal, it is necessary and sufficient that the slopes of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

are $2 \times 1/2$. But the first Newton slope is

$$\min\left\{1/2,\ \operatorname{ord}(a^{\sigma^g} + d\frac{c^{\sigma^g}}{c})\right\}.$$

Therefore, a is not a unit.

Conversely, given such a matrix, we just define the Frobenius on

$$\bigoplus_{i \in \mathbb{Z}/g\mathbb{Z}} \mathbb{D}_i,$$

where each $\mathbb{D}_i$ is simply a rank-2 free $W(k)$-module, by using the matrices

$$\begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

in the obvious way.

There is a unique way to define V, and we get a supersingular crystal together with an admissible basis. The case of g even is similar. □

Remark 5.4.2. If we change the original basis by a matrix M as in (5.2), then the resulting matrix Q' for Frobenius is given by

$$Q' = M^{-1} Q M^{\sigma^g}.$$

If we put

$$Q_1 = \begin{cases} \begin{pmatrix} a & b \\ c & d \end{pmatrix}, & g \equiv 1 \pmod 2, \\ \begin{pmatrix} b & -pa \\ d & -pc \end{pmatrix}, & g \equiv 0 \pmod 2, \end{cases}$$

and

$$Q' = \begin{cases} (-1)^{(g-1)/2} p^{(g-1)/2} Q_1', & g \equiv 1 \pmod 2, \\ (-1)^{(g-2)/2} p^{(g-2)/2} Q_1', & g \equiv 0 \pmod 2, \end{cases}$$

then

$$Q_1' = M^{-1} Q' M^{\sigma^g}.$$

Example 5.4.3. Let us consider the case

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix},$$

which gives a superspecial crystal for every g. Here are some examples of the universal deformation of $\mathbb{D}_g$:

g	deformation matrix
1	$\begin{pmatrix} S_1 & -p \\ 1 & 0 \end{pmatrix}$
2	$\begin{pmatrix} S_1S_2 - p & -pS_1 \\ S_2 & -p \end{pmatrix}$
3	$\begin{pmatrix} S_1S_2S_3 - p(S_1 + S_3) & -pS_1S_2 + p^2 \\ S_2S_3 - p & -pS_2 \end{pmatrix}$
4	$\begin{pmatrix} S_1S_2S_3S_4 - p(S_1S_2 + S_1S_4 + S_3S_4) + p^2 & -pS_1S_2S_3 + p^2(S_1 + S_3) \\ S_2S_3S_4 - p(S_2 + S_4) & -pS_2S_3 + p^2 \end{pmatrix}$
5	$\begin{pmatrix} S_1S_2S_3S_4S_5 - p(S_1S_2S_5 + S_1S_4S_5 + S_1S_2S_3 + S_3S_4S_5) + p^2(S_1 + S_3 + S_5) & -pS_1S_2S_3S_4 + p^2(S_1S_2 + S_1S_4 + S_3S_4) - p^3 \\ S_2S_3S_4S_5 - p(S_2S_5 + S_4S_5 + S_2S_3) + p^2 & -pS_2S_3S_4 + p^2(S_2 + S_4) \end{pmatrix}$

Here $S_1 = T_g, S_2 = T^{\sigma}_{g-1}$, etc.

Theorem 5.4.4. *For every g, for every totally real field L of degree g over $\mathbb{Q}$, and for every algebraically closed field k of characteristic p, where p is inert in L, there exists a unique superspecial ppRM crystal.*

Proof. As explained above, existence follows from Corollary 2.3.12. We prove uniqueness.

Consider first the case where g is odd. We consider the matrix

$$Q_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

as defining a σ^g-crystal structure on $W(k) \oplus W(k)$, which we denote by $(\mathbb{H}, \Phi)$. Put $\Sigma = \sigma^g$. We claim that there exists a matrix $N \in GL_2(W(k))$ such that

$$N^{-1}Q_1N^{\Sigma} = \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}. \tag{5.4}$$

Indeed, Q_1 defines a rank-two supersingular Σ-crystal, and hence, if we let $\mathbb{H}_{1/2} = W(k)[F]/(F^2+p)$ with F, by definition, Σ-linear, then we have an injection

$$\mathbb{H} \hookrightarrow \mathbb{H}_{1/2},$$

because of the semi-simplicity of the category of Σ-crystals $\otimes\mathbb{Q}$.

We claim that, in fact, $(\mathbb{H}, \Phi)$ is isomorphic to $(\mathbb{H}_{1/2}, F)$. To see that, we argue as follows. Let

$$P = \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}.$$

We may think of $(\mathbb{H}_{1/2}, F)$ as $(W(k)^2, P)$, where P is a linear operator, that is, we consider

$$\mathbb{H}_{1/2} \underset{W(k)^{\Sigma}}{\otimes} W(k).$$

Under this interpretation, the image of $\mathbb{H}$ is a lattice $\mathcal{L}$ of $W(k)^2$ that is stable under P.

Choose $x \in \mathcal{L} \setminus P\mathcal{L}$. Let $y = Px$. Then $\{x, y\}$ is a basis of $\mathcal{L}$ over $W(k)$ such that the matrix describing P with respect to this basis is P again. It follows that $(\mathcal{L}, P|_{\mathcal{L}})$ is isomorphic to $(W(k)^2, P)$. Thus,

$$(\mathbb{H}, \Phi) \cong (\mathbb{H}_{1/2}, F)$$

as Σ-crystals. Hence, there exists a basis $Y = \{y_1, y_2\}$ of $\mathbb{H}$ such that Φ is given by

$$\begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}$$

with respect to this basis. Let N be the change of basis matrix.

One immediately checks that for $N_1 \in GL_2(W(k))$ we have

$$N_1^{-1} \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix} N_1^{\Sigma} = \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix} \tag{5.5}$$

if and only if $N_1 \in GL_2(W(\mathbb{F}_{p^{2g}}))$ and is of the form

$$\begin{pmatrix} n_1 & -pn_3^{\Sigma} \\ n_3 & n_1^{\Sigma} \end{pmatrix}.$$

On the other hand, equation (5.4) gives

$$\det(N)^{\Sigma} = \det(N).$$

Therefore, since the norm map

$$\text{Norm} : W(\mathbb{F}_{p^{2g}})^{\times} \longrightarrow W(\mathbb{F}_{p^{g}})^{\times}$$

is surjective, we can find N_1 satisfying equation (5.5) (e.g. by taking $n_3 = 0$) such that $\det(NN_1) = 1$. Let $M = NN_1 1$; then

$$M^{-1}Q_1M^{\Sigma} = \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}$$

and $\det(M) = 1$. Thus M defines a symplectic change of coordinates of $\mathbb{D}_1$, automatically satisfying the condition (5.3).

The case of even g is quite similar. In this case, the matrix

$$Q_1 = \begin{pmatrix} b & -pa \\ d & -pc \end{pmatrix}$$

is divisible by p. Put $Q_1 = -pR$ for some $R \in SL_2(W(k))$. We think of the matrix R as defining a Σ-crystal of rank 2 whose slopes are both equal to zero. It is therefore, by a theorem of Katz ([Ka], Theorem 1.6.1), split over an algebraically closed field. Therefore, we see that there exists a matrix $N \in GL_2(W(k))$ such that

$$N^{-1}RN = I_2.$$

The same argument as before shows we may choose $N \in SL_2(W(k))$. By multiplying by a suitable lower or upper triangular matrix we can in fact get N to also satisfy equation (5.3). Therefore, N gives us the desired change of basis for $\mathbb{D}_1$. □

Theorem 5.4.5. *Assume that $p > 2$. Let $\beta_0 < \beta_1 < \cdots < \beta_s$ be the possible Newton polygons as given in Theorem 5.2.1. Let x be a geometric superspecial point of $\mathcal{M}_n$. For every ℓ, $0 \leq \ell \leq s$, there are g irreducible nonsingular formal subschemes $V_1^\ell, \cdots, V_g^\ell$ of the formal neighborhood of x such that the following hold.*

1. *The type of the generic point of V_i^ℓ is i and the Newton polygon is β_ℓ.*
2. *Every geometric deformation with type i and Newton polygon β_ℓ is a point of V_i^ℓ.*

The equations defining V_i^ℓ are the following:

$$t_i = 0,\ t_{i-1}^{p^2} + t_{i+1} = 0, \cdots,\ t_{i-\ell+1}^{p^{2\ell-2}} + t_{i+\ell-1} = 0.$$

Proof. Using Theorem 5.4.4, we may assume that the crystal $(\mathbb{D}, F)$ defined as $(\mathbb{D}(X_x(p)), F)$ is given by

$$\left(\bigoplus_{i=1}^{g} W(k),\ \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}\right).$$

Consider the universal deformation of x of type i. To ease notation we will assume that $i = g$, and hence $t_g = 0$. We compute the effect of the universal deformation, given by a matrix A, on the Σ-crystal $\mathbb{D}_{(g-1)/2}$ if g is odd, and on $\mathbb{D}_{g/2}$ if g is even.

g odd: We have

$$A = \cdots \begin{pmatrix} T_2^{\sigma^{x-1}} & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} T_1^{\sigma^x} & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} T_{g-1}^{\sigma^{x+2}} & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} T_{g-2}^{\sigma^{x+3}} & -p \\ 1 & 0 \end{pmatrix} \cdots,$$

where $x = (g-3)/2$. Multiplying the middle five matrices, we get

$$A = -p \times \cdots \begin{pmatrix} T_3^{\sigma^{x-2}} & -p \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} (T_1 + T_{g-1}^{\sigma^2})^{\sigma^x} T_2^{\sigma^{x-1}} T_{g-2}^{\sigma^{x+3}} - pT_2^{\sigma^{x-1}} & -p(T_1 + T_{g-1}^{\sigma^2})^{\sigma^x} T_2^{\sigma^{x-1}} + p^2 \\ (T_1 + T_{g-1}^{\sigma^2})^{\sigma^x} T_{g-2}^{\sigma^{x+3}} - p & -p(T_1 + T_{g-1}^{\sigma^2})^{\sigma^x} \end{pmatrix} \times \begin{pmatrix} T_{g-3}^{\sigma^{x+4}} & -p \\ 1 & 0 \end{pmatrix} \cdots.$$

Some reflection shows that

$$A = -p \times \begin{pmatrix} \alpha \times u_1 + p \times * & -p\alpha \times u_2 + p^2 \times * \\ * & p \times * \end{pmatrix},$$

where

$$\alpha = (T_1 + T_{g-1}^{\sigma^2})^{\sigma^x}$$

and u_1, u_2 are a product of Teichmüller lifts of the t_i for $i \neq g$ raised to a certain power. Applying the formula for the first Newton slope, we get that the Newton polygon is weakly above β_1 iff

$$p | \alpha\ u_1 \alpha^{\sigma^g} u_2^{\sigma^g}.$$

In other words,

$$p | (T_1 + T_{g-1}^{\sigma^2}).$$

However,

$$T_1 + T_{g-1}^{\sigma^2} = (t_1, 0, 0, \cdots) + (t_{g-1}^{p^2}, 0, 0, \cdots) = (t_1 + t_{g-1}^{p^2}, \cdots),$$

and we get in particular

$$t_1 = -t_{g-1}^{p^2}.$$

If $p > 2$ then -1 is the Teichmüller lift of -1, and hence we get from the multiplicativity of the Teichmüller map that in fact $T_1 + T_{g-1}^{p^2} = 0$.

We go back to the product defining A, and we get

$$A = -p \times \cdots \begin{pmatrix} T_2^{\sigma^{x-1}} & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} T_{g-2}^{\sigma^{x+3}} & -p \\ 1 & 0 \end{pmatrix} \cdots$$

and therefore we may argue in the same way.

g even: The story is very similar. The same kind of computations leave us with

$$A = -p \times \begin{pmatrix} T_{g/2} & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \times u_1 + p \times * & -p\alpha \times u_2 + p^2 \times * \\ * & p \times * \end{pmatrix},$$

which gives

$$A = -p \times \begin{pmatrix} \alpha \times u_1' + p \times * & -p\alpha \times u_2' + p^2 \times * \\ * & p \times * \end{pmatrix},$$

and the same arguments apply. □

Remark 5.4.6. The proof clearly shows that the theorem holds for $p = 2$ as well, but the equations are more complicated due to the fact that the Teichmüller lift of -1 is not -1 in this case.

Corollary 5.4.7. *The weak Grothendieck conjecture holds for principally polarized abelian varieties with real multiplication.*

Proof. Theorem 5.4.5 gives, for every $0 \leq i \leq s$, the existence of a geometric point t of $\mathcal{M}_n$ with $a(X_t) = 1$ and $\mathcal{N}_t = \beta_i$. Hence, the corollary follows from Theorem 5.3.3. □

The Newton polygon.

We determine the Newton polygon of the generic point of every component of W_τ. As before, we denote by

$$\beta_0 < \beta_1 \cdots < \beta_s, \quad s = \left[\frac{g+1}{2}\right],$$

the possible Newton polygons.

Definition 5.4.8. A subset ρ of $\mathbb{Z}/g\mathbb{Z}$ is called *spaced* if $i \in \rho \Rightarrow i+1 \notin \rho$. Given $\tau \subseteq \mathbb{Z}/g\mathbb{Z}$, we let

$$\lambda(\tau) = \max\{|\rho| : \rho \subseteq \tau, \rho \text{ spaced}\}$$

(except for g odd and $\tau = \{1, \cdots, g\}$, where we put $\lambda(\tau) = (g+1)/2$).

Definition 5.4.9. Given a type τ, we may write it uniquely as

$$\tau = K_1 \coprod \cdots \coprod K_k, \tag{5.6}$$

where each K_i is a maximal set of consecutive elements of τ (i.e., the blocks of τ). We put $\mathrm{par}(K_i) = 1$ if K_i is of odd length, and $\mathrm{par}(K_i) = 0$ if K_i is of even length.

The following lemma is elementary.

Lemma 5.4.10. *For every type τ,*

$$\lambda(\tau) = \frac{1}{2}\sum_{i=1}^{k} |K_i| + \mathrm{par}(K_i). \quad \square$$

Theorem 5.4.11. *The Newton polygon of the generic point of every component of W_τ is $\beta_{\lambda(\tau)}$.*

Proof. Using Remark 2.3.10, Corollary 2.3.6 and Corollary 2.3.12, we reduce to studying the question locally at a superspecial point. By Theorem 5.4.4 we may further assume that we are in the following situation:

Consider the standard superspecial crystal $(\mathbb{D}, \Phi)$ as in Example 5.4.3. It is displayed by

$$\begin{pmatrix} \mathfrak{d}_2(0,0,\cdots,0) & \mathfrak{d}_2(-1,-1,\cdots,-1) \\ \mathfrak{d}_2(1,1,\cdots,1) & \mathfrak{d}_2(0,0,\cdots,0) \end{pmatrix}.$$

The action of the Frobenius of the universal deformation on $\mathbb{D}_g$ is given by

$$\begin{pmatrix} S_1 & -p \\ 1 & 0 \end{pmatrix}\begin{pmatrix} S_2 & -p \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} S_g & -p \\ 1 & 0 \end{pmatrix},$$

where $S_1 = T_g, S_2 = T^\sigma_{g-1}$, etc. It will be convenient to use the notation

$$M_i(S_i) = \begin{pmatrix} S_i & -p \\ 1 & 0 \end{pmatrix}.$$

Let

$$M = M_1(S_1)\cdots M_g(S_g).$$

We shall prove that if we put $S_i = 0$ for $i \in \tau$, then the first Newton slope of M is $\lambda(\tau)$. This would conclude the proof. We do it in two steps.

Step 1. *We can assume that τ is spaced.* Indeed, write τ in blocks,

$$\tau = K_1 \coprod \cdots \coprod K_k,$$

as in equation (5.6). If $\mathrm{par}(K_i) = 0$, substitute 0 for any S_j in the block K_i. If $\mathrm{par}(K_i) = 1$, substitute 0 for any S_j in the block K_i that is not the first element in the block. That is, if $K_i = n_i \cdots m_i$, then we put

$$S_{n_i+1} = \cdots = S_{m_i} = 0.$$

For $\mathrm{par}(K_i) = 0$, this substitution has the effect

$$\prod_{j=n_i}^{m_i} M_i(S_i) = (-p)^{(m_i-n_i-1)/2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For $\mathrm{par}(K_i) = 1$, it has the effect

$$\prod_{j=n_i}^{m_i} M_i(S_i) = (-p)^{(m_i-n_i)/2}\begin{pmatrix} S_{n_i} & -p \\ 1 & 0 \end{pmatrix}.$$

We can pull those powers of p outside the product $M_1(S_1)\cdots M_g(S_g)$ that we are trying to understand. What we are left with is a spaced type. That is, those n_i for i's such that $\mathrm{par}(K_i) = 1$ are a spaced type in $\mathbb{Z}/g\mathbb{Z}$ with the indices of all the S_i's we equated to zero deleted. Using Lemma 5.4.10, we see that we may assume that τ is spaced.

Step 2. *τ is spaced.* Without loss of generality, we assume that the first index in τ is 1 (otherwise compute on a suitable $\mathbb{D}_i$ instead of $\mathbb{D}_g$). Put

$$\tau = \{1 < i_2 < \cdots < i_a\}.$$

Note that $i_a < g$ and that $\lambda(\tau) = a$. Put

$$\tau_j = \tau \cap \{1, \cdots, j\}.$$

Let

$$\epsilon_i = \begin{cases} 0, & i \in \tau, \\ 1, & i \notin \tau. \end{cases}$$

We prove, by induction on j in the set

$$\mathcal{I} = \{2, 3, \cdots, i_2 - 1, i_2 + 1, \cdots, i_3 - 1, i_3 + 1, \cdots, g\},$$

that the ordered product

$$\prod_{i=1}^{j}\begin{pmatrix} \epsilon_i S_i & -p \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p^{|\tau_j|}a_j & p^{|\tau_j|+1}b_j \\ p^{|\tau_j|-1}c_j & p^{|\tau_j|}d_j \end{pmatrix}, \tag{5.7}$$

where

$$\mathrm{ord}_p(a_j) = \mathrm{ord}_p(c_j) = \mathrm{ord}_p(d_j) = 0. \tag{5.8}$$

First, for $j = 2$ we get

$$\begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix}\begin{pmatrix} S_2 & -p \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -p & 0 \\ S_2 & -p \end{pmatrix},$$

and the induction starts. In the induction step we distinguish cases:

(i) We have $j \in \mathcal{I}$ and $j + 1 \in \mathcal{I}$. Then

$$\begin{pmatrix} p^{|\tau_j|}a_j & p^{|\tau_j|+1}b_j \\ p^{|\tau_j|-1}c_j & p^{|\tau_j|}d_j \end{pmatrix}\begin{pmatrix} S_{j+1} & -p \\ 1 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} p^{|\tau_j|}a_j S_{j+1} + p^{|\tau_j|+1}b_j & p^{|\tau_j|+1}\cdot(-a_j) \\ p^{|\tau_j|-1}c_j S_{j+1} + p^{|\tau_j|}d_j & p^{|\tau_j|}c_j \end{pmatrix}.$$

Note that $|\tau_j| = |\tau_{j+1}|$, and the induction step is proved.

(ii) We have $j \in \mathcal{I}$ and $j+1 \notin \mathcal{I}$. Then,

$$\begin{pmatrix} p^{|\tau_j|}a_j p^{|\tau_j|+1}b_j & \\ p^{|\tau_j|-1}c_j & p^{|\tau_j|}d_j \end{pmatrix} \begin{pmatrix} 0 & -p \\ 1 & 0 \end{pmatrix} \begin{pmatrix} S_{j+2} & -p \\ 1 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} p^{|\tau_j|}a_j & p^{|\tau_j|+1}b_j \\ p^{|\tau_j|-1}c_j & p^{|\tau_j|}d_j \end{pmatrix} \begin{pmatrix} -p & 0 \\ S_{j+2} & -p \end{pmatrix}$$
$$= \begin{pmatrix} p^{|\tau_j|+1}\cdot(-a_j) + p^{|\tau_j|+1}b_jS_{j+2} & p^{|\tau_j|+2}\cdot(-b_j) \\ p^{|\tau_j|}\cdot(-c_j) + p^{|\tau_j|}d_jS_{j+2} & p^{|\tau_j|+1}\cdot(-d_j) \end{pmatrix}$$
$$= \begin{pmatrix} p^{|\tau_{j+2}|}(b_jS_{j+2} - a_j) & p^{|\tau_{j+2}|+1}\cdot(-b_j) \\ p^{|\tau_{j+2}|-1}(d_jS_{j+2} - c_j) & p^{|\tau_{j+2}|}\cdot(-d_j) \end{pmatrix}.$$

The induction step follows easily. Using equations (5.7) and (5.8), as well as Lemma 5.3.1, we see that the first Newton slope is $\lambda(\tau)$. □

Corollary 5.4.12. *Every isogeny class of p-divisible groups appearing in Theorem 5.2.1 is realized for some point of $\mathcal{M}_n$.* □

Remark 5.4.13. One could also prove this directly using Honda-Tate theory.

Corollary 5.4.14. *For every i, some component of $\mathcal{N}_i$ is of dimension greater than or equal to $g - i$.*

Proof. Indeed, we just need to find a type τ such that

$$|\tau| = \lambda(\tau) = i.$$

Take for example

$$\tau = \{1, 3, \cdots, 2i+1\}.$$

(For g odd, $i = (g+1)/2$, use Theorem 5.4.5.) □

Finally, our methods in this paper suggest that the following conjecture, if true, would reduce Conjecture 5.2.5 to a fairly explicit local question.

Conjecture 5.4.15. *Every component of every Newton polygon strata contains a superspecial point.*

References

[BG1] Bachmat, E. and Goren, E. Z.: On the non ordinary locus in Hilbert-Blumenthal surfaces, *Math. Ann.* **313** (1999), 475–506.

[CN] Chai, C. L. and Norman, P.: Bad reduction of the Siegel moduli scheme of genus two with $\Gamma_0(p)$-level structure, *Amer. J. of Math.* **112** (1990), 1003-1071.

[Dm] Demazure, M.: *Lectures on p-divisible groups*, Lecture Notes in Math. **302**, Springer-Verlag, 2nd printing, 1986.

[DP] Deligne, P. and Pappas, G.: Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant, *Compositio Math.* **90** (1994), 59 - 74.

[EO] Ekedahl, T. and Oort, F. : Connected subsets of a moduli space of abelian varieties (*to appear*).

[F] Fulton W: *Intersection Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **3** Folge, Band **2**, Springer-Verlag, Berlin, Heidelberg, 1984.

[vdG] van der Geer, G.: *Hilbert modular surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **3** Folge, Band **16**, Springer-Verlag, Berlin, Heidelberg, 1988.

[vdG1] van der Geer, G.: Cycles on the moduli space of abelian varieties, preprint **96-10**, Dept. of Math. Kyoto University, 1996.

[G] Goren, E. Z.: Hasse invariants for Hilbert modular varieties, CICMA preprint 1998-10, submitted, 16 pp.

[Gr] Grothendieck, A.: *Groupes de Barsotti-Tate et cristaux de Dieudonné*, Séminaire de Mathématiques Supérieures **45**, Presses de l'Université de Montréal, 1974.

[Ka] Katz, N. M.: Slope filtration of F-crystals, *Astérisque* **63** (1979), 113-164.

[K2] Katz, N. M.: Serre-Tate local moduli, in: *Surfaces algébriques*, Sém. Géom. Alg. Orsay 1976-78 (Eds. J. Giraud, L. Illusie, and M. Raynaud), Lecture Notes in Math. **868**, Springer-Verlag, New York, 1981, pp. 138-202.

[K] Koblitz, N.: p-adic variation of the zeta-function over families of varieties defined over finite fields, *Compositio Math.* **31** (1975), 119-218.

[KO] Kraft, H. P. and Oort, F.: *In preparation.*

[Me] Messing, W.: *The crystals associated to Barsotti-Tate groups: with applications to abelian schemes*, Lecture Notes in Math. **264**, Springer-Verlag, New York, 1972.

[MB] Moret-Bailly, L.: Pinceaux de variétés abéliennes, *Astérisque* **129**, 1985.

[Mu] Mumford, D.: Towards an enumerative geometry of the moduli space of curves, *Geometry and Arithmetic II*, Birkhäuser, 1983, pp. 271–328.

[N] Norman, P.: An algorithm for computing local moduli of abelian varieties, *Annals of Math.* **101** (1975), 499 - 509.

[NO] Norman, P. and Oort, F.: Moduli of abelian varieties, *Annals of Math.* **112** (1980), 413-439.

[O1] Oort, F.: A stratification of a moduli space of polarized abelian varieties in positive characteristic, preprint nr. **997**, Dept. of Math. Utrecht University, 1997.

[O2] Oort, F.: Moduli of abelian varieties, finite group schemes and formal groups, preprint nr. **1038**, Dept. of Math. Utrecht University, 1997.

[O3] Oort, F.: Newton polygons and formal groups: conjectures by Manin and Grothendieck, preprint nr. **995**, Dept. of Math. Utrecht University, 1997.

[O4] Oort, F.: Moduli of abelian varieties in positive characteristic, In: Barsotti symposium in Algebraic Geometry, 253 - 275, Perspectives in Math. **15**, Academic Press, 1994, pp. 253–276.

[Ra] Rapoport, M.; On the bad reduction of Shimura varieties, *Automorphic forms, Shimura varieties, and L-functions, Vol. II* (Ann Arbor, MI, 1988), Perspectives in Math., vol. 11, Academic Press, 1990, pp. 253–321.

[Z] Zink, T.: The display of a formal p-divisible group, Universität Bielefeld, Preprint **98-017**, February 1998, 155 pp.

Department of Mathematics and Statistics, 805 Sherbrooke St. W., McGill University, Montreal, Quebec, Canada, H3A 2K6
E-mail address: goren@math.mcgill.ca

Mathematisch Instituut, Budapestlaan 6, Utrecht University, NL - 3508 TA Utrecht, The Netherlands
E-mail address: oort@math.uu.nl