



# COMPOSITIO MATHEMATICA

## Intersection theory on Shimura surfaces

Benjamin Howard

Compositio Math. **145** (2009), 423–475.

[doi:10.1112/S0010437X09003935](https://doi.org/10.1112/S0010437X09003935)



FOUNDATION  
COMPOSITIO  
MATHEMATICA

*The London  
Mathematical  
Society*





# Intersection theory on Shimura surfaces

Benjamin Howard

## ABSTRACT

Kudla has proposed a general program to relate arithmetic intersection multiplicities of special cycles on Shimura varieties to Fourier coefficients of Eisenstein series. The lowest dimensional case, in which one intersects two codimension one cycles on the integral model of a Shimura curve, has been completed by Kudla, Rapoport and Yang. In the present paper we prove results in a higher dimensional setting. On the integral model of a Shimura surface we consider the intersection of a Shimura curve with a codimension two cycle of complex multiplication points, and relate the intersection to certain cycle classes constructed by Kudla, Rapoport and Yang. As a corollary we deduce that our intersection multiplicities appear as Fourier coefficients of a Hilbert modular form of half-integral weight.

## 1. Introduction

Suppose that  $B_0$  is an indefinite quaternion division algebra over  $\mathbb{Q}$ . Let  $G_0$  be the algebraic group over  $\mathbb{Q}$  whose functor of points is  $G_0(A) = (B_0 \otimes_{\mathbb{Q}} A)^\times$  for any  $\mathbb{Q}$ -algebra  $A$ , fix a maximal order  $\mathcal{O}_{B_0} \subset B_0$ , and set  $U_0^{\max} = \widehat{\mathcal{O}}_{B_0}^\times$  and

$$\Gamma_0^{\max} = G_0(\mathbb{Q}) \cap U_0^{\max}.$$

After fixing an isomorphism  $G_0(\mathbb{R}) \cong \mathrm{GL}_2(\mathbb{R})$  the group  $G_0(\mathbb{R})$  acts on the complex manifold  $X_0 = \mathbb{C} \setminus \mathbb{R}$ , and the quotient

$$\mathcal{M}_0(\mathbb{C}) = \Gamma_0^{\max} \backslash X_0$$

is isomorphic to the complex points of a Shimura curve over  $\mathbb{Q}$  which parametrizes abelian surfaces over  $\mathbb{Q}$ -schemes with an action of  $\mathcal{O}_{B_0}$ . Extending the moduli problem over  $\mathrm{Spec}(\mathbb{Z})$  one obtains an algebraic (Deligne–Mumford) stack  $\mathcal{M}_0$ . Let  $\mathrm{Sym}_2(\mathbb{Z})$  denote the  $\mathbb{Z}$ -module of symmetric  $2 \times 2$  matrices with entries in  $\mathbb{Z}$  and set

$$\mathrm{Sym}_2(\mathbb{Z})^\vee = \left\{ \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \mid a, b, c \in \mathbb{Z} \right\}.$$

For every  $T \in \mathrm{Sym}_2(\mathbb{Z})^\vee$  Kudla, Rapoport and Yang [KRY06] have constructed an arithmetic cycle class

$$\widehat{\mathcal{Z}}(T, \mathbf{v}) \in \widehat{\mathrm{CH}}_{\mathbb{R}}^2(\mathcal{M}_0)$$

in the Gillet–Soulé arithmetic Chow group (with real coefficients and modified for stacks, as in [KRY06, ch. 2]) of the arithmetic surface  $\mathcal{M}_0$ . Here  $\mathbf{v} \in M_2(\mathbb{R})$  is a symmetric positive definite

---

Received 30 January 2008, accepted in final form 8 September 2008, published online 25 March 2009.

*2000 Mathematics Subject Classification* 14G35, 14G40, 11F41.

*Keywords:* Shimura varieties, Arakelov theory, Kudla’s program.

This research was supported in part by NSF grant DMS-0556174, and by a Sloan Foundation Research Fellowship.

This journal is © Foundation Compositio Mathematica 2009.

parameter. Letting  $\tau = \mathbf{u} + i\mathbf{v} \in \mathfrak{h}_2$  denote the variable on the Siegel half-space of genus two, these classes have the remarkable property that the generating series

$$\widehat{\phi}_2(\tau) = \sum_{T \in \text{Sym}_2(\mathbb{Z})^\vee} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \cdot q^T$$

is a non-holomorphic Siegel modular form of weight  $3/2$ . Here the isomorphism

$$\widehat{\text{deg}} : \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M}_0) \rightarrow \mathbb{R}$$

is the *arithmetic degree* of [KRY06, (2.4.10)]. Pulling  $\widehat{\phi}_2$  back to the product of two upper half-planes via the diagonal embedding  $\mathfrak{h}_1 \times \mathfrak{h}_1 \rightarrow \mathfrak{h}_2$  results in a modular form of parallel weight  $3/2$  for a congruence subgroup of  $\text{SL}_2(\mathbb{Z}) \times \text{SL}_2(\mathbb{Z})$ , and Kudla, Rapoport and Yang express the Fourier coefficients of the pullback in terms of the arithmetic intersections of classes in  $\widehat{\text{CH}}_{\mathbb{R}}^1(\mathcal{M}_0)$ . More precisely, for each  $t \in \mathbb{Z}$  and  $v \in \mathbb{R}^+$  they define a class

$$\widehat{\mathcal{Z}}(t, v) \in \widehat{\text{CH}}_{\mathbb{R}}^1(\mathcal{M}_0)$$

and prove that the pullback of  $\widehat{\phi}_2$  by the diagonal embedding has Fourier expansion

$$\widehat{\phi}_2(\tau_1, \tau_2) = \sum_{t_1, t_2 \in \mathbb{Z}} \langle \widehat{\mathcal{Z}}(t_1, v_1), \widehat{\mathcal{Z}}(t_2, v_2) \rangle q_1^{t_1} q_2^{t_2} \tag{1}$$

where  $\tau_j = u_j + iv_j$  is a variable in the upper half-plane  $\mathfrak{h}_1$ ,  $q_j = e^{2\pi i \tau_j}$ , and the pairing  $\langle \cdot, \cdot \rangle$  is the Gillet–Soulé intersection pairing

$$\widehat{\text{CH}}_{\mathbb{R}}^1(\mathcal{M}_0) \times \widehat{\text{CH}}_{\mathbb{R}}^1(\mathcal{M}_0) \rightarrow \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M}_0) \xrightarrow{\widehat{\text{deg}}} \mathbb{R}.$$

The proof of (1) amounts to proving the decomposition [KRY06, Theorem C]

$$\langle \widehat{\mathcal{Z}}(t_1, v_1), \widehat{\mathcal{Z}}(t_2, v_2) \rangle = \sum_T \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \tag{2}$$

in which the sum is over all  $T \in \text{Sym}_2(\mathbb{Z})^\vee$  of the form  $T = \begin{pmatrix} t_1 & * \\ * & t_2 \end{pmatrix}$ , and  $\mathbf{v}$  is the diagonal matrix with diagonal entries  $v_1, v_2$ .

The question which motivates this paper is the following. Given a real quadratic field  $F \subset \mathbb{R}$  and a  $\mathbb{Z}$ -basis  $\{\varpi_1, \varpi_2\}$  of  $\mathcal{O}_F$  (which we now fix once and for all) one can define, as in Equation (50), a twisted embedding  $\mathfrak{h}_1 \times \mathfrak{h}_1 \rightarrow \mathfrak{h}_2$  in such a way that the pullback of  $\widehat{\phi}_2$  to  $\mathfrak{h}_1 \times \mathfrak{h}_1$  is a half-integral weight Hilbert modular form. Can one interpret the Fourier coefficients of the twisted pullback as arithmetic intersection multiplicities of cycles on a Shimura variety? To answer this, let

$$B = B_0 \otimes_{\mathbb{Q}} F, \quad \mathcal{O}_B = \mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathcal{O}_F.$$

Throughout this paper we assume the following equivalent conditions on  $F$  and  $B_0$ .

HYPOTHESIS A. We assume that:

- (a)  $\text{disc}(B_0) \cdot \mathcal{O}_F = \text{disc}(B)$ ;
- (b)  $\mathcal{O}_B$  is a maximal order of  $B$ ;
- (c) every prime divisor of  $\text{disc}(B_0)$  splits in  $F$ .

Let  $G$  be the algebraic group over  $\mathbb{Q}$  whose functor of points is

$$G(A) = \{g \in (B \otimes_{\mathbb{Q}} A)^\times \mid \text{Nm}(g) \in A^\times\}$$

for any  $\mathbb{Q}$ -algebra  $A$ , where  $\text{Nm}$  denotes the reduced norm on  $B$ . Define a maximal compact open subgroup  $U^{\max} \subset G(\mathbb{A}_f)$  by

$$U^{\max} = \{g \in \widehat{\mathcal{O}}_B^\times \mid \text{Nm}(g) \in \widehat{\mathbb{Z}}^\times\}$$

and set

$$\Gamma^{\max} = G(\mathbb{Q}) \cap U^{\max}.$$

Let  $\sigma$  be the nontrivial Galois automorphism of  $F/\mathbb{Q}$  and identify

$$F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R} \times \mathbb{R} \tag{3}$$

using the map  $x \otimes 1 \mapsto (x, x^\sigma)$ . The earlier choice of isomorphism  $G_0(\mathbb{R}) \cong \text{GL}_2(\mathbb{R})$  together with (3) determine an isomorphism

$$G(\mathbb{R}) \cong \{(x, y) \in \text{GL}_2(\mathbb{R}) \times \text{GL}_2(\mathbb{R}) \mid \det(x) = \det(y)\}$$

in such a way that the subgroup  $G_0(\mathbb{R}) \subset G(\mathbb{R})$  is identified with the diagonal. Let  $X \subset X_0 \times X_0$  be the subset of pairs whose imaginary parts are either both positive or both negative. The compact complex manifold

$$\mathcal{M}(\mathbb{C}) = \Gamma^{\max} \backslash X$$

is then identified with the complex points of a Shimura surface which parametrizes abelian fourfolds over  $\mathbb{Q}$ -schemes with an action of  $\mathcal{O}_B$  (and some additional polarization data which we ignore in this introduction). The obvious embedding  $\mathcal{M}_0(\mathbb{C}) \rightarrow \mathcal{M}(\mathbb{C})$  induced by the inclusion of  $X_0$  into  $X$  has a moduli theoretic meaning: an abelian surface  $A_0$  with  $\mathcal{O}_{B_0}$  action is taken to the abelian fourfold  $A = A_0 \otimes \mathcal{O}_F$  (Serre’s tensor construction [Con04, § 7]) with its induced action of  $\mathcal{O}_B = \mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathcal{O}_F$ . Extending the moduli problems over  $\text{Spec}(\mathbb{Z})$  we obtain a closed immersion  $\mathcal{M}_0 \rightarrow \mathcal{M}$  of an arithmetic surface into an arithmetic threefold.

On the other hand, the complex surface  $\mathcal{M}(\mathbb{C})$  comes equipped with a natural family of cycles of dimension zero. For each totally positive  $\alpha \in \mathcal{O}_F$  we consider the finite set  $\mathcal{Y}(\alpha)(\mathbb{C})$  of isomorphism classes of abelian fourfolds over  $\mathbb{C}$  equipped with commuting actions of  $\mathcal{O}_B$  and  $\mathcal{O}_F[\sqrt{-\alpha}]$ . There is an evident function

$$\mathcal{Y}(\alpha)(\mathbb{C}) \rightarrow \mathcal{M}(\mathbb{C})$$

which forgets the  $\mathcal{O}_F[\sqrt{-\alpha}]$ -action. Extending the moduli problem across  $\text{Spec}(\mathbb{Z})$  one obtains a finite morphism  $\mathcal{Y}(\alpha) \rightarrow \mathcal{M}$  in which  $\mathcal{Y}(\alpha)$  is an algebraic stack over  $\mathbb{Z}$  of dimension at most two, and which has dimension one after restricting to  $\mathbb{Z}[\text{disc}(B_0)^{-1}]$ . At a prime  $p$  dividing  $\text{disc}(B_0)$ , the stack  $\mathcal{Y}(\alpha)_{/\mathbb{Z}_p}$  may (depending on  $\alpha$ ) have vertical components of dimension two. Our central object of study is a certain class

$$\widehat{\mathcal{Y}}(\alpha, v) \in \widehat{\text{CH}}^2(\mathcal{M})$$

in the codimension two Gillet–Soulé arithmetic Chow group of  $\mathcal{M}$  (with rational coefficients) whose construction is based on the moduli problem  $\mathcal{Y}(\alpha)$ . Here  $v \in F \otimes_{\mathbb{Q}} \mathbb{R}$  is an auxiliary totally positive parameter. This class is obtained by first modifying the vertical components of  $\mathcal{Y}(\alpha)$  at the prime divisors of  $\text{disc}(B_0)$  to obtain a cycle class in the codimension two Chow group  $\text{CH}_{\mathcal{Y}(\alpha)}^2(\mathcal{M})$  with support on  $\mathcal{Y}(\alpha)$ , and then augmenting this cycle with a Green current of the type constructed by Kudla [Kud97] and Kudla, Rapoport and Yang [KRY06]. It is the definition of the cycle class, and especially of the construction of the vertical components at primes dividing  $\text{disc}(B_0)$  carried out in § 4, which is the primary original contribution of this work.

The closed immersion  $\mathcal{M}_0 \rightarrow \mathcal{M}$  induces the  $\mathbb{Q}$ -linear functional

$$\widehat{\text{deg}}_{\mathcal{M}_0} : \widehat{\text{CH}}^2(\mathcal{M}) \rightarrow \mathbb{R}$$

called the *arithmetic degree along  $\mathcal{M}_0$* . Our main result, which appears in the text as Theorem 5.1.4, is a twisted form of the decomposition (2).

**THEOREM B.** *Suppose that  $\alpha \in \mathcal{O}_F$  and  $v \in F \otimes_{\mathbb{Q}} \mathbb{R}$  are totally positive and that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic. Then*

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\mathcal{Y}}(\alpha, v) = \sum_{T \in \Sigma(\alpha)} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v})$$

where  $v = (v_1, v_2) \in \mathbb{R} \times \mathbb{R}$  and  $\mathbf{v}$  are related by (48). The set  $\Sigma(\alpha)$  appearing in the sum is defined as

$$\Sigma(\alpha) = \left\{ \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \in \text{Sym}_2(\mathbb{Z})^\vee \mid \alpha = a\varpi_1^2 + b\varpi_1\varpi_2 + c\varpi_2^2 \right\} \tag{4}$$

where  $\{\varpi_1, \varpi_2\}$  is the  $\mathbb{Z}$ -basis of  $\mathcal{O}_F$  fixed above, and used in the definition of the embedding  $\mathfrak{h}_1 \times \mathfrak{h}_1 \rightarrow \mathfrak{h}_2$  of (50).

The assumption that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic in the theorem is made to ensure that

$$(\mathcal{Y}(\alpha) \times_{\mathcal{M}} \mathcal{M}_0)_{/\mathbb{Q}} = \emptyset,$$

i.e. that  $\mathcal{Y}(\alpha)$  and  $\mathcal{M}_0$  do not meet in the generic fiber. The theorem should remain true without this assumption; see [How] for results in this direction. As a corollary of the theorem we deduce in § 5.2 the following partial generalization of (1).

**COROLLARY C.** *The pullback of  $\widehat{\phi}_2$  via the twisted embedding  $\mathfrak{h}_1 \times \mathfrak{h}_1 \rightarrow \mathfrak{h}_2$  has a Fourier expansion of the form*

$$\widehat{\phi}_2(\tau_1, \tau_2) = \sum_{\alpha \in \mathcal{O}_F} c(\alpha, v) \cdot q^\alpha$$

where  $v = (v_1, v_2)$  is the imaginary part of  $(\tau_1, \tau_2)$  and  $q^\alpha = e^{2\pi i \tau_1 \alpha} e^{2\pi i \tau_2 \alpha^{\sigma}}$ . If  $\alpha$  is totally positive and  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic then

$$c(\alpha, v) = \widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\mathcal{Y}}(\alpha, v).$$

Allowing ourselves the luxury of speculation, we now explain how these results fit into the program of generalized Gross–Zagier style theorems proposed by Kudla [Kud04]. Suppose first that one can extend the definition of the arithmetic class  $\widehat{\mathcal{Y}}(\alpha, v)$  to all  $\alpha \in \mathcal{O}_F$ , as opposed to restricting only to totally positive  $\alpha$ , in such a way that (passing to arithmetic Chow groups with real coefficients) the generating series

$$\widehat{\theta}(\tau_1, \tau_2) = \sum_{\alpha \in \mathcal{O}_F} \widehat{\mathcal{Y}}(\alpha, v) \cdot q^\alpha \in \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M})[[q]]$$

is a non-holomorphic vector-valued Hilbert modular form of parallel weight  $3/2$  for the congruence subgroup (51). If Theorem B were extended to every  $\alpha \in \mathcal{O}_F$  then (after extending the arithmetic degree along  $\mathcal{M}_0$  to arithmetic Chow groups with real coefficients) we would have the equality of  $q$ -expansions

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\theta}(\tau_1, \tau_2) = \widehat{\phi}_2(\tau_1, \tau_2).$$

Given another Hilbert modular form  $f$  of parallel weight  $3/2$  for (51), now with real coefficients, one could then define the *arithmetic theta lift* of  $f$  as the Petersson inner product

$$\widehat{\Theta}(f) \stackrel{\text{def}}{=} \langle f(\tau_1, \tau_2), \widehat{\theta}(\tau_1, \tau_2) \rangle^{\text{Pet}} \in \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M}),$$

and then

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\Theta}(f) = \langle f(\tau_1, \tau_2), \widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\theta}(\tau_1, \tau_2) \rangle^{\text{Pet}} = \langle f(\tau_1, \tau_2), \widehat{\phi}_2(\tau_1, \tau_2) \rangle^{\text{Pet}}.$$

On the automorphic side, one may attach to  $f$  and  $B_0$  a Rankin  $L$ -function obtained by integrating  $f$  against the twisted pullback of the genus two Eisenstein series  $\mathcal{E}_2(\tau, s, B_0)$  studied by Kudla, Rapoport and Yang (see § 5.2),

$$L(f, s, B_0) = \langle f(\tau_1, \tau_2), \mathcal{E}_2(\tau_1, \tau_2, s, B_0) \rangle^{\text{Pet}}.$$

The Eisenstein series vanishes at  $s = 0$ , and a fundamental result of Kudla, Rapoport and Yang [KRY06, Theorem B] is that

$$\mathcal{E}'_2(\tau, 0, B_0) = \widehat{\phi}_2(\tau).$$

Pulling back this equality to  $\mathfrak{h}_1 \times \mathfrak{h}_1$ , it follows that

$$\langle f(\tau_1, \tau_2), \widehat{\phi}_2(\tau_1, \tau_2) \rangle^{\text{Pet}} = L'(f, 0, B_0).$$

Thus Theorem B may be viewed as a first step toward the Gross–Zagier style result

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\Theta}(f) = L'(f, 0, B_0).$$

The  $L$ -function  $L(f, s, B_0)$  is somewhat mysterious. One would like to know, for example, if it admits an Euler product, and if so what form the Euler factors take. We hope to address this question in a subsequent article.

### 1.1 Notation

Throughout this paper the symbols  $F$ ,  $\sigma$ ,  $G$ , and  $G_0$  have the same meanings as above,  $\mathfrak{D}$  denotes the different of  $F/\mathbb{Q}$ , Hypothesis A is assumed, and the  $\mathbb{Z}$ -basis  $\{\varpi_1, \varpi_2\}$  of  $\mathcal{O}_F$  is fixed. We choose the maximal order  $\mathcal{O}_{B_0} \subset B_0$  to be stable under the main involution and, as on [BC92, p. 127], choose an  $s \in \mathcal{O}_B$  such that  $s^2 = -\text{disc}(B_0)$ . Define a positive involution of  $B_0$ ,

$$b \mapsto b^* = s^{-1} \cdot b^t \cdot s.$$

Extend  $b \mapsto b^*$  to a positive involution of  $B$ , trivial on  $F$ . Let  $\mathbb{A}_f$  denote the ring of finite adeles of  $\mathbb{Q}$ .

## 2. Arithmetic intersection theory

Our basic references for stacks are [Gil84, LM00], and [Vis89]. By an *algebraic stack* we always mean a Deligne–Mumford stack in the sense of [LM00]. If  $S$  is a scheme and  $\mathcal{X}$  is an algebraic stack then we denote by  $\mathcal{X}(S)$  the fiber of  $\mathcal{X}$  over  $S$ . The goal of § 2 is to develop a rudimentary (and somewhat *ad-hoc*) extension to stacks of the Gillet–Soulé arithmetic intersection theory [GS91, Sou92].

### 2.1 Chow groups

Throughout all of § 2.1 we fix an algebraic stack  $\mathcal{M}$  separated and of finite type over a regular Noetherian scheme  $S$ . We describe below the basics of the theory of Chow groups of such a stack.

If we assume that  $\mathcal{M}$  is a scheme, then the theory is described in [Ful98] and [Sou92]. If instead we assume that  $S$  is the spectrum of a field, then the theory is described in [Gil84] and [Vis89], but for the very rudimentary results described below the methods extend easily to the case of  $S$  regular and Noetherian.

An *irreducible cycle* on  $\mathcal{M}$  is a nonempty integral closed substack of  $\mathcal{M}$ . For any nonnegative  $k \in \mathbb{Z}$  let  $Z^k(\mathcal{M})$  be the free  $\mathbb{Q}$ -module generated by the codimension  $k$  irreducible cycles on  $\mathcal{M}$ . Elements of  $Z^k(\mathcal{M})$  are called *cycles* of codimension  $k$ . Define the group of *rational equivalences* in codimension  $k$ ,

$$R^k(\mathcal{M}) = \bigoplus_{\mathcal{D}} k(\mathcal{D})^\times \tag{5}$$

where the direct sum is over all irreducible cycles  $\mathcal{D}$  of codimension  $k - 1$  on  $\mathcal{M}$ , and  $k(\mathcal{D})$  is the field of rational functions on  $\mathcal{D}$  (i.e. the quotient field of [Vis89, Definition 1.14]). There is a  $\mathbb{Z}$ -module map

$$\partial : R^k(\mathcal{M}) \rightarrow Z^k(\mathcal{M})$$

which takes  $f \in k(\mathcal{D})$  to its Weil divisor viewed as a cycle on  $\mathcal{M}$  (defined for schemes in [Ful98] or [Sou92] and extended to algebraic stacks as in [Gil84, § 4.4]). Define the *codimension  $k$  Chow group* (with rational coefficients)

$$\text{CH}^k(\mathcal{M}) = \text{coker}(R^k(\mathcal{M}) \otimes_{\mathbb{Z}} \mathbb{Q} \xrightarrow{\partial} Z^k(\mathcal{M})).$$

There is also a notion of Chow group with support along a closed subscheme, or, slightly more generally, with support along a proper map. Suppose we are given an algebraic stack  $\mathcal{Y}$  over  $S$  and a proper map  $\phi : \mathcal{Y} \rightarrow \mathcal{M}$ . Let  $Z_{\mathcal{Y}}^k(\mathcal{M}) \subset Z^k(\mathcal{M})$  be the subspace generated by the irreducible cycles supported on the image of  $\mathcal{Y}$ . Similarly, let  $R_{\mathcal{Y}}^k(\mathcal{M}) \subset R^k(\mathcal{M})$  be defined exactly as in (5), but where the direct sum is over only those irreducible cycles  $\mathcal{D}$  supported on the image of  $\mathcal{Y}$ . The map  $\partial$  defined above restricts to a map  $\partial : R_{\mathcal{Y}}^k(\mathcal{M}) \rightarrow Z_{\mathcal{Y}}^k(\mathcal{M})$ , and the *codimension  $k$  Chow group with support along  $\mathcal{Y}$*  (again, with rational coefficients) is defined as

$$\text{CH}_{\mathcal{Y}}^k(\mathcal{M}) = \text{coker}(R_{\mathcal{Y}}^k(\mathcal{M}) \otimes_{\mathbb{Z}} \mathbb{Q} \xrightarrow{\partial} Z_{\mathcal{Y}}^k(\mathcal{M})).$$

Suppose that  $\mathcal{Y}$  and  $\mathcal{M}_0$  are algebraic stacks and that there is a finite type *flat* morphism  $f : \mathcal{M}_0 \rightarrow \mathcal{M}$  and a proper morphism  $\mathcal{Y} \rightarrow \mathcal{M}$ . As in [Gil84, Proposition 4.6(i)] or [Vis89, Proposition 3.7] there is an induced *flat pullback* on Chow groups

$$f^* : \text{CH}_{\mathcal{Y}}^k(\mathcal{M}) \rightarrow \text{CH}_{\mathcal{Y}_0}^k(\mathcal{M}_0)$$

in which  $\mathcal{Y}_0 = \mathcal{Y} \times_{\mathcal{M}} \mathcal{M}_0$ .

**DEFINITION 2.1.1.** Let  $T$  be a regular Noetherian scheme over  $S$ . A  *$T$ -uniformization* of  $\mathcal{M}$  is an isomorphism  $\mathcal{M}_{/T} \cong [H \backslash M]$  of stacks in which  $M$  is a scheme over  $T$  and  $H$  is a finite group of automorphisms of  $M$ .

**LEMMA 2.1.2.** *Suppose we have a  $T$ -uniformization  $\mathcal{M}_{/T} \cong [H \backslash M]$  and let  $f : M \rightarrow \mathcal{M}_{/T}$  be the canonical morphism. If  $\mathcal{Y} \rightarrow \mathcal{M}$  is any proper morphism of algebraic stacks and  $Y = \mathcal{Y} \times_{\mathcal{M}} M$ , then the flat pullback*

$$f^* : \text{CH}_{\mathcal{Y}_{/T}}^k(\mathcal{M}_{/T}) \rightarrow \text{CH}_Y^k(M)^H \tag{6}$$

is an isomorphism.

*Proof.* As in [Gil84, Proposition 4.6(iii)] or [Vis89, Proposition 3.7] there is a push-forward homomorphism  $f_* : \text{CH}_Y^k(\mathcal{M}) \rightarrow \text{CH}_{\mathcal{Y}/T}^k(\mathcal{M}/_T)$  which satisfies

$$(f_* \circ f^*)(\mathcal{C}) = |H| \cdot \mathcal{C}$$

and

$$(f^* \circ f_*)(\mathcal{C}) = \sum_{h \in H} h \cdot \mathcal{C}.$$

Therefore, the flat pullback (6) has inverse  $|H|^{-1} \cdot f_*$ . □

Suppose  $S = \text{Spec}(R)$  with  $R$  a discrete valuation ring. Let  $\eta$  and  $s$  be the generic point and closed point of  $S$ , respectively, and assume that  $\mathcal{M}$  is proper and flat over  $S$ . An irreducible cycle  $\mathcal{C}$  on  $\mathcal{M}$  is *horizontal* if it is flat over  $S$ , and is *vertical* if it is supported on the special fiber  $\mathcal{M}/_{k(s)}$ . Let

$$\text{CH}_{\text{ver}}^k(\mathcal{M}) \stackrel{\text{def}}{=} \text{CH}_{\mathcal{M}/_{k(s)}}^k(\mathcal{M})$$

denote the Chow group with support along the special fiber  $\mathcal{M}/_{k(s)} \rightarrow \mathcal{M}$ , and define  $Z_{\text{ver}}^k(\mathcal{M})$  in the same way. If  $\mathcal{Y} \rightarrow \mathcal{M}$  is any morphism of algebraic stacks there is a canonical homomorphism

$$Z_{\mathcal{Y}}^k(\mathcal{M}) \rightarrow Z_{\text{ver}}^k(\mathcal{M}) \oplus Z_{\mathcal{Y}/_{k(\eta)}}^k(\mathcal{M}/_{k(\eta)})$$

which takes an irreducible cycle  $\mathcal{C}$  to  $(\mathcal{C}, 0)$  if  $\mathcal{C}$  is vertical, and to  $(0, \mathcal{C}/_{k(\eta)})$  if  $\mathcal{C}$  is horizontal. If the image of  $\mathcal{Y}/_{k(\eta)} \rightarrow \mathcal{M}/_{k(\eta)}$  has codimension at least  $k$  (including possibly  $\mathcal{Y}/_{k(\eta)} = \emptyset$ ) then this homomorphism descends to a map

$$\text{CH}_{\mathcal{Y}}^k(\mathcal{M}) \rightarrow \text{CH}_{\text{ver}}^k(\mathcal{M}) \oplus Z_{\mathcal{Y}/_{k(\eta)}}^k(\mathcal{M}/_{k(\eta)}) \tag{7}$$

as in [Sou92, Remark III.2.1].

### 2.2 A little K-theory

For any algebraic stack  $\mathcal{M}$  we let  $\mathbf{K}_0(\mathcal{M})$  denote the Grothendieck group of the category of coherent  $\mathcal{O}_{\mathcal{M}}$ -modules (denoted by  $K'_0(\mathcal{M})$  in [GS87] and [Sou92]). That is, the free abelian group generated by coherent  $\mathcal{O}_{\mathcal{M}}$ -modules, modulo the subgroup generated by the relations  $\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2$  whenever there is an exact sequence  $0 \rightarrow \mathcal{F}_1 \rightarrow \mathcal{F} \rightarrow \mathcal{F}_2 \rightarrow 0$ . The class in  $\mathbf{K}_0(\mathcal{M})$  of a coherent sheaf  $\mathcal{F}$  will be denoted  $[\mathcal{F}]$ . If  $\phi : \mathcal{Y} \rightarrow \mathcal{M}$  is a proper morphism of algebraic stacks we define  $\mathbf{K}_0^{\mathcal{Y}}(\mathcal{M})$  to be the Grothendieck group of the category of coherent  $\mathcal{O}_{\mathcal{M}}$ -modules supported on the image of  $\mathcal{Y}$ . There is an obvious homomorphism  $\mathbf{K}_0^{\mathcal{Y}}(\mathcal{M}) \rightarrow \mathbf{K}_0(\mathcal{M})$  (which is typically not injective). There is a *higher direct image* map

$$R\phi_* : \mathbf{K}_0(\mathcal{Y}) \rightarrow \mathbf{K}_0^{\mathcal{Y}}(\mathcal{M}) \tag{8}$$

defined by

$$R\phi_*[\mathcal{F}] = \sum_{k \geq 0} (-1)^k [R^k \phi_* \mathcal{F}].$$

If  $\phi$  is a closed immersion or a finite map, so that  $\phi_*$  is an exact functor from coherent  $\mathcal{O}_{\mathcal{Y}}$ -modules to coherent  $\mathcal{O}_{\mathcal{M}}$ -modules, then  $R\phi_*[\mathcal{F}] = [\phi_* \mathcal{F}]$ . If  $\mathcal{Z} \rightarrow \mathcal{M}$  is a proper map with image contained in the image of  $\mathcal{Y} \rightarrow \mathcal{M}$  then there is an evident change of support map  $\mathbf{K}_0^{\mathcal{Z}}(\mathcal{M}) \rightarrow \mathbf{K}_0^{\mathcal{Y}}(\mathcal{M})$ . Define a decreasing filtration on  $\mathbf{K}_0^{\mathcal{Y}}(\mathcal{M})$ ,

$$F^k \mathbf{K}_0^{\mathcal{Y}}(\mathcal{M}) = \bigcup_{\mathcal{Z}} \text{image}(\mathbf{K}_0^{\mathcal{Z}}(\mathcal{M}) \rightarrow \mathbf{K}_0^{\mathcal{Y}}(\mathcal{M})),$$



where the union is over the closed substacks  $\mathcal{Z} \rightarrow \mathcal{M}$  of codimension at least  $k$  which are contained in the image of  $\mathcal{Y}$ .

Let  $M$  be a regular scheme which is separated and of finite type over a regular Noetherian base  $S$ , and let  $Y \rightarrow M$  be a proper map of schemes. Define  $K_0^Y(M)$  to be the free group generated by quasi-isomorphism classes of finite complexes  $\mathcal{P}^k \rightarrow \dots \rightarrow \mathcal{P}^0$  of coherent locally free  $\mathcal{O}_M$ -modules whose homology sheaves  $H^i(\mathcal{P}^\bullet)$  are supported on the image of  $Y$ , modulo the subgroup generated by the relations  $\mathcal{P}^\bullet = \mathcal{Q}_1^\bullet + \mathcal{Q}_2^\bullet$  if there is an exact sequence  $0 \rightarrow \mathcal{Q}_1^\bullet \rightarrow \mathcal{P}^\bullet \rightarrow \mathcal{Q}_2^\bullet \rightarrow 0$ . The class in  $K_0^Y(M)$  of a complex  $\mathcal{P}^\bullet$  is denoted  $[\mathcal{P}^\bullet]$ . If  $\mathcal{F}$  is a coherent sheaf on  $M$  supported on the image of  $Y$  then by [GS87, § 1.9] there is a finite resolution

$$\mathcal{P}^\bullet \rightarrow \mathcal{F} \rightarrow 0$$

of  $\mathcal{F}$  by coherent locally free  $\mathcal{O}_M$ -modules. The rule  $[\mathcal{F}] \mapsto [\mathcal{P}^\bullet]$  then defines an isomorphism  $\mathbf{K}_0^Y(M) \cong K_0^Y(M)$  with inverse  $[\mathcal{P}^\bullet] \mapsto \sum_i (-1)^i H^i(\mathcal{P}^\bullet)$ . In particular, the group  $K_0^Y(M)$  inherits a filtration  $F^k K_0^Y(M)$  from the filtration on  $\mathbf{K}_0^Y(M)$  defined above. Suppose that  $i : M_0 \rightarrow M$  is a finite type morphism of schemes with  $M_0$  regular and separated over  $S$ , and set  $Y_0 = Y \times_M M_0$ . If  $i$  is flat then the functor  $i^*$  from coherent  $\mathcal{O}_M$ -modules to coherent  $\mathcal{O}_{M_0}$ -modules is exact, and so the rule  $[\mathcal{F}] \mapsto [i^* \mathcal{F}]$  is a well-defined homomorphism  $i^* : \mathbf{K}_0^Y(M) \rightarrow \mathbf{K}_0^{Y_0}(M_0)$ . The groups  $K_0^Y(M)$  admit pullback maps even when  $i$  is not flat: according to [Sou92, Theorem I.3(iii)] the rule  $[\mathcal{P}^\bullet] \mapsto [i^* \mathcal{P}^\bullet]$  defines a homomorphism

$$i^* : F^k K_0^Y(M) \rightarrow F^k K_0^{Y_0}(M_0).$$

Keeping  $M$  and  $Y$  as in the preceding paragraph, if  $Z \rightarrow M$  is an irreducible cycle of codimension  $k$  supported on  $Y$  then we may define  $\alpha(Z)$  to be the image of  $[\mathcal{O}_Z]$  under

$$\mathbf{K}_0(Z) \rightarrow \mathbf{K}_0^Z(M) \rightarrow F^k \mathbf{K}_0^Y(M) \cong F^k K_0^Y(M).$$

By a theorem of Gillet–Soulé, see [GS87, Theorem 8.2] or [Sou92, § 3.3], the rule  $Z \mapsto \alpha(Z)$  defines an isomorphism

$$\mathrm{CH}_Y^k(M) \cong G_r^k K_0^Y(M) \otimes_{\mathbb{Z}} \mathbb{Q}, \tag{9}$$

where  $G_r^k K_0^Y(M) = F^k K_0^Y(M) / F^{k+1} K_0^Y(M)$ . Combining this with the pullback on K-theory constructed above, we obtain a pullback homomorphism on Chow groups,

$$i^* : \mathrm{CH}_Y^k(M) \rightarrow \mathrm{CH}_{Y_0}^k(M_0). \tag{10}$$

If  $i$  is flat then this homomorphism is the flat pullback constructed in § 2.1.

LEMMA 2.2.1. *Let  $S$  be a Noetherian scheme and let  $\eta \in S$  be the generic point of an irreducible component  $D$  of  $S$ . We give  $D$  its reduced subscheme structure and view  $\mathcal{O}_D$  as a quotient sheaf of  $\mathcal{O}_S$ . For any coherent sheaf  $\mathcal{F}$  on  $S$  there is a closed subscheme  $Z \rightarrow S$  not containing  $D$  such that*

$$[\mathcal{F}] - \mathrm{length}_{\mathcal{O}_{S,\eta}}(\mathcal{F}_\eta) \cdot [\mathcal{O}_D] \in \mathrm{Im}(\mathbf{K}_0(Z) \rightarrow \mathbf{K}_0(S)).$$

*Proof.* Let  $U$  be an open affine neighborhood of  $\eta$ . By [Mat89, Theorem 6.5(iii)], any coherent  $\mathcal{O}_U$ -module  $\mathcal{A}$  which satisfies  $\mathcal{A}_\eta \neq 0$  contains a subsheaf isomorphic to  $\mathcal{O}_D|_U$ . Taking  $\mathcal{A} = \mathcal{F}|_U$  and using induction on  $\mathrm{length}_{\mathcal{O}_{S,\eta}}(\mathcal{F}_\eta)$ , we find a coherent  $\mathcal{O}_U$ -module  $\mathcal{G}$  with trivial stalk at  $\eta$  and satisfying

$$[\mathcal{F}|_U] = \mathrm{length}_{\mathcal{O}_{S,\eta}}(\mathcal{F}_\eta) \cdot [\mathcal{O}_D|_U] + [\mathcal{G}]$$

in  $\mathbf{K}_0(U)$ . But  $\mathcal{G}|_V = 0$  for some open neighborhood  $V \subset U$  of  $\eta$ . Thus  $[\mathcal{F}] - \text{length}_{\mathcal{O}_{S,\eta}}(\mathcal{F}_\eta) \cdot [\mathcal{O}_D]$  lies in the kernel of restriction  $\mathbf{K}_0(S) \rightarrow \mathbf{K}_0(V)$ , and hence is contained in the image of  $\mathbf{K}_0(S \setminus V) \rightarrow \mathbf{K}_0(S)$  by the exact sequence of [Sou92, Lemma I.3.2].  $\square$

LEMMA 2.2.2. *Let  $S$  be a Noetherian scheme and let  $\Omega$  denote any subset of the set of irreducible components of  $S$ . There is an exact sequence*

$$\bigoplus_Z \mathbf{K}_0(Z) \rightarrow \mathbf{K}_0(S) \rightarrow \bigoplus_{D \in \Omega} \mathbf{K}_0(\text{Spec}(\mathcal{O}_{S,\eta})) \rightarrow 0$$

in which the first sum is over all closed subschemes of  $S$  having no components contained in  $\Omega$ , and in the second sum  $\eta$  is the generic point of the component  $D$ .

*Proof.* By induction it suffices to treat the case in which  $\Omega$  consists of a single irreducible component of  $S$  with generic point  $\eta$ . As  $\mathcal{O}_{S,\eta}$  is Artinian there is a canonical isomorphism

$$\mathbf{K}_0(\text{Spec}(\mathcal{O}_{S,\eta})) \cong \mathbb{Z}$$

given by  $[\mathcal{F}_\eta] \mapsto \text{length}_{\mathcal{O}_{S,\eta}}(\mathcal{F}_\eta)$ , and so the claim follows easily from Lemma 2.2.1.  $\square$

### 2.3 Arithmetic Chow groups

DEFINITION 2.3.1. An *arithmetic stack* is an algebraic stack  $\mathcal{M}$  over  $\mathbb{Z}$  satisfying the following properties:

- (a)  $\mathcal{M}$  is regular;
- (b) the structure map  $\mathcal{M} \rightarrow \text{Spec}(\mathbb{Z})$  is flat and projective;
- (c) for every prime  $\ell$  one can find a positive integer  $N$  which is not divisible by  $\ell$  and a  $\mathbb{Z}[1/N]$ -uniformization  $\mathcal{M}_{/\mathbb{Z}[1/N]} \cong [H \setminus M]$  in the sense of Definition 2.1.1.

Throughout all of § 2.3 we work with a fixed arithmetic stack  $\mathcal{M}$ , equidimensional of dimension  $d = \dim(\mathcal{M})$ , and choose a  $\mathbb{Q}$ -uniformization  $\mathcal{M}_{/\mathbb{Q}} \cong [H_{\mathbb{Q}} \setminus M_{\mathbb{Q}}]$ . Thus  $M_{\mathbb{Q}}$  is a smooth projective variety over  $\mathbb{Q}$  of dimension  $d - 1$ . There is a notion of a Green current [Sou92, § III.1] for a cycle in  $Z^k(M_{\mathbb{Q}})$ , and as a consequence of [Gil84, § 4] there are canonical isomorphisms

$$Z^k(\mathcal{M}_{/\mathbb{Q}}) \cong Z^k(M_{\mathbb{Q}})^{H_{\mathbb{Q}}}, \quad R^k(\mathcal{M}_{/\mathbb{Q}}) \cong R^k(M_{\mathbb{Q}})^{H_{\mathbb{Q}}}.$$

Hence we may define a Green current for a cycle  $\mathcal{C}_{\mathbb{Q}} \in Z^k(\mathcal{M}_{/\mathbb{Q}})$  to be an  $H_{\mathbb{Q}}$ -invariant Green current for the corresponding cycle  $C_{\mathbb{Q}} \in Z^k(M_{\mathbb{Q}})^{H_{\mathbb{Q}}}$ . This definition does not depend on the choice of  $\mathbb{Q}$ -uniformization: if  $\mathcal{M}_{/\mathbb{Q}} \cong [H'_{\mathbb{Q}} \setminus M'_{\mathbb{Q}}]$  is another  $\mathbb{Q}$ -uniformization then one may form a third  $\mathbb{Q}$ -uniformization,

$$\mathcal{M}_{/\mathbb{Q}} \cong [(H_{\mathbb{Q}} \times H'_{\mathbb{Q}}) \setminus (M_{\mathbb{Q}} \times_{\mathcal{M}_{/\mathbb{Q}}} M'_{\mathbb{Q}})],$$

which allows one to identify  $H_{\mathbb{Q}}$ -invariant currents on  $M_{\mathbb{Q}}$  with  $H'_{\mathbb{Q}}$ -invariant currents on  $M'_{\mathbb{Q}}$ , as both are identified with  $H_{\mathbb{Q}} \times H'_{\mathbb{Q}}$ -invariant currents on  $M_{\mathbb{Q}} \times_{\mathcal{M}_{/\mathbb{Q}}} M'_{\mathbb{Q}}$ . Using the notation of [Sou92, § III.1], if  $\mathcal{C} \in Z^k(\mathcal{M})$  and  $\Xi, \Xi' \in D^{k-1, k-1}(M_{\mathbb{Q}})$  are two Green currents for  $\mathcal{C}_{/\mathbb{Q}}$ , we will say that  $\Xi$  and  $\Xi'$  are *equivalent* if there are  $H_{\mathbb{Q}}$ -invariant currents  $u \in D^{k-2, k-1}(M_{\mathbb{Q}})$  and  $v \in D^{k-1, k-2}(M_{\mathbb{Q}})$  such that  $\Xi - \Xi' = \partial(u) + \bar{\partial}(v)$ .

Let  $\widehat{Z}^k(\mathcal{M})$  be the  $\mathbb{Q}$ -vector space of pairs  $(\mathcal{C}, \Xi)$  in which  $\mathcal{C} \in Z^k(\mathcal{M})$  and  $\Xi$  is an equivalence class of Green currents for  $\mathcal{C}_{/\mathbb{Q}}$ . For each irreducible cycle  $D \in Z^{k-1}(M_{\mathbb{Q}})$  and  $f \in k(D)^\times$  there is an associated Green current, constructed in [Sou92, § III.1] and denoted  $[-\log |f|^2]$ , for the Weil

divisor  $\text{div}(f) \in Z^k(M_{\mathbb{Q}})$ . Thus every element of  $R^k(M_{\mathbb{Q}})$  has a canonically associated Green current. Taking  $H_{\mathbb{Q}}$ -invariants yields a  $\mathbb{Z}$ -module map

$$\widehat{\partial} : R^k(\mathcal{M}) \rightarrow \widehat{Z}^k(\mathcal{M}),$$

and we define the *codimension  $k$  arithmetic Chow group* (with rational coefficients) by

$$\widehat{\text{CH}}^k(\mathcal{M}) = \text{coker}(R^k(\mathcal{M}) \otimes_{\mathbb{Z}} \mathbb{Q} \xrightarrow{\widehat{\partial}} \widehat{Z}^k(\mathcal{M})).$$

For every prime  $p$  there is a homomorphism

$$\text{CH}_{\text{ver}}^k(\mathcal{M}/\mathbb{Z}_p) \rightarrow \widehat{\text{CH}}^k(\mathcal{M}) \tag{11}$$

defined by endowing a cycle on  $\mathcal{M}/\mathbb{F}_p$  with the trivial Green current.

Fix a prime  $p$ . For any irreducible cycle  $\mathcal{C}$  on  $\mathcal{M}/\mathbb{Z}_p$  of codimension  $d$  (necessarily vertical) define

$$\text{deg}_p(\mathcal{C}) = \sum_{x \in \mathcal{C}(\mathbb{F}_p^{\text{alg}})} \frac{1}{|\text{Aut}_{\mathcal{M}}(x)|}$$

where the sum is over all isomorphism classes of objects in the category  $\mathcal{C}(\mathbb{F}_p^{\text{alg}})$ . By  $\text{Aut}_{\mathcal{M}}(x)$  we mean the automorphism group of  $x$  in the category  $\mathcal{M}(\mathbb{F}_p^{\text{alg}})$ . Extending the definition of  $\text{deg}_p$  linearly to all of  $Z_{\text{ver}}^d(\mathcal{M}/\mathbb{Z}_p)$ , there is an induced homomorphism

$$\text{deg}_p : \text{CH}_{\text{ver}}^d(\mathcal{M}/\mathbb{Z}_p) \rightarrow \mathbb{Q}.$$

Suppose  $(\mathcal{C}, \Xi) \in \widehat{Z}^d(\mathcal{M})$ . By definition  $\Xi$  is a  $(d - 1, d - 1)$ -current (up to equivalence) on the compact complex manifold  $M_{\mathbb{Q}}(\mathbb{C})$  of dimension  $d - 1$ . Thus we may define

$$\text{deg}_{\infty}(\Xi) = \frac{1}{2 \cdot |H|} \int_{M_{\mathbb{Q}}(\mathbb{C})} \Xi$$

where the integral means evaluation of the current  $\Xi$  at the constant function 1. The function on  $\widehat{Z}^d(\mathcal{M})$  defined by

$$\widehat{\text{deg}}(\mathcal{C}, \Xi) = \text{deg}_{\infty}(\Xi) + \sum_p \text{deg}_p(\mathcal{C}/\mathbb{Z}_p) \cdot \log(p)$$

then descends to a  $\mathbb{Q}$ -linear map, the *arithmetic degree*,

$$\widehat{\text{deg}} : \widehat{\text{CH}}^d(\mathcal{M}) \rightarrow \mathbb{R}.$$

Indeed, it suffices to check that if  $\mathcal{W}$  is an irreducible cycle of dimension one on  $\mathcal{M}$  and  $f$  is a rational function on  $\mathcal{W}$ , then the arithmetic degree of the pair  $(\text{div}(f), [-\log |f|^2])$  is 0. If  $\mathcal{W}$  is a horizontal cycle then this is precisely the calculation of [KRY06, (2.1.11)]. If  $\mathcal{W}$  is a vertical cycle supported in characteristic  $p$ , then one chooses a  $\mathbb{Z}_p$ -uniformization  $[H \setminus M] \cong \mathcal{M}/\mathbb{Z}_p$ , sets  $W = M \times_{\mathcal{M}} \mathcal{W}$ , and uses the fact that principal Weil divisors on  $W$  have degree 0.

Now suppose that  $\mathcal{M}_0$  is another arithmetic stack, equidimensional of dimension  $d_0 = \dim(\mathcal{M}_0)$ , and that  $i : \mathcal{M}_0 \rightarrow \mathcal{M}$  is a closed immersion. We will define a  $\mathbb{Q}$ -linear functional, the *arithmetic degree along  $\mathcal{M}_0$* ,

$$\widehat{\text{deg}}_{\mathcal{M}_0} : \widehat{\text{CH}}^{d_0}(\mathcal{M}) \rightarrow \mathbb{R}.$$

Morally speaking, the arithmetic degree along  $\mathcal{M}_0$  should be the composition

$$\widehat{\text{CH}}^{d_0}(\mathcal{M}) \xrightarrow{i^*} \widehat{\text{CH}}^{d_0}(\mathcal{M}_0) \xrightarrow{\widehat{\text{deg}}} \mathbb{R}$$

of the pullback  $i^*$  as constructed in [GS91, § 4.4] and [Sou92, § III.3] with the arithmetic degree defined above. This requires extending the theory of pullbacks in [GS91] from schemes to arithmetic stacks, and is further complicated by a error in the construction of  $i^*$  identified and corrected by Gubler [Gub02]. Presumably this extension can be successfully done, but we will avoid the issue by borrowing a trick from Bruinier, Burgos and Kühn [BBK07]. Suppose we have an integer  $N$  and a  $\mathbb{Z}[1/N]$ -uniformization  $\mathcal{M}_{/\mathbb{Z}[1/N]} \cong [H \backslash M]$ . Set  $M_0 = \mathcal{M}_0 \times_{\mathcal{M}} M$ . There is an obvious homomorphism

$$\delta_{M/\mathcal{M}} : \widehat{\text{CH}}^{d_0}(\mathcal{M}) \rightarrow \widehat{\text{CH}}^{d_0}(M)$$

induced by pullback of cycles through the composition  $M \rightarrow \mathcal{M}_{/\mathbb{Z}[1/N]} \rightarrow \mathcal{M}$ , and the full force of the theories of [GS91] and [BKK07] (both of which include arithmetic Chow groups of schemes over  $\mathbb{Z}[1/N]$ ) may be applied to the arithmetic Chow group  $\widehat{\text{CH}}^{d_0}(M)$ . In particular, there is a homomorphism of  $\mathbb{Q}$ -vector spaces

$$\Delta_N : \widehat{\text{CH}}^{d_0}(\mathcal{M}) \rightarrow \widehat{\text{CH}}^1(\mathbb{Z}[1/N])$$

obtained as the composition

$$\widehat{\text{CH}}^{d_0}(\mathcal{M}) \xrightarrow{\delta_{M/\mathcal{M}}} \widehat{\text{CH}}^{d_0}(M) \xrightarrow{|H|^{-1}} \widehat{\text{CH}}^{d_0}(M) \rightarrow \widehat{\text{CH}}^{d_0}(M_0) \rightarrow \widehat{\text{CH}}^1(\mathbb{Z}[1/N])$$

in which the third arrow is the pullback of [GS91, § 4.4] and the final arrow is the push-forward of [GS91, § 3.6]. The map  $\Delta_N$  depends on  $N$  but not on the choice of  $\mathbb{Z}[1/N]$ -uniformization  $M$ . If we define  $\mathbb{R}_N$  to be the quotient of  $\mathbb{R}$  by the additive subgroup generated of all rational linear combinations of  $\{\log(p)\}$ , as  $p$  ranges over the prime divisors of  $N$ , then there is a canonical  $\mathbb{Q}$ -linear map

$$\widehat{\text{deg}}_{\mathbb{Z}[1/N]} : \widehat{\text{CH}}^1(\mathbb{Z}[1/N]) \rightarrow \mathbb{R}_N$$

obtained by imitating the construction of the arithmetic degree  $\widehat{\text{CH}}^1(\mathbb{Z}) \rightarrow \mathbb{R}$ . As in [BBK07, § 6.3] there is a canonical isomorphism

$$\mathbb{R} \cong \varprojlim_{N \in \mathbb{Z}^+} \mathbb{R}_N,$$

and our definition of an arithmetic stack guarantees that one can find a coterminal family  $\mathcal{N} \subset \mathbb{Z}^+$  of this inverse system such that for every  $N \in \mathcal{N}$  the stack  $\mathcal{M}$  admits a  $\mathbb{Z}[1/N]$ -uniformization. This allows us to define

$$\widehat{\text{deg}}_{\mathcal{M}_0} : \widehat{\text{CH}}^{d_0}(\mathcal{M}) \rightarrow \mathbb{R}$$

by

$$\widehat{\text{deg}}_{\mathcal{M}_0} = \varprojlim_{N \in \mathcal{N}} \widehat{\text{deg}}_{\mathbb{Z}[1/N]} \circ \Delta_N.$$

Now suppose  $\mathcal{Y} \rightarrow \mathcal{M}$  is a proper morphism of algebraic stacks such that the image of  $\mathcal{Y}_{/\mathbb{Q}} \rightarrow \mathcal{M}_{/\mathbb{Q}}$  has codimension  $k$ . If  $p$  is a prime, by a *local cycle datum at  $p$*  of codimension  $k$  with support on  $\mathcal{Y}$  we mean a triple  $(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_p)$  in which

$$\mathcal{C}_{\mathbb{Q}} \in Z_{\mathcal{Y}_{/\mathbb{Q}}}^k(\mathcal{M}_{/\mathbb{Q}}),$$

$\Xi$  is an equivalence class of Green currents for  $\mathcal{C}_{\mathbb{Q}}$ , and

$$\mathcal{C}_p \in \text{CH}_{\mathcal{Y}_{/\mathbb{Z}_p}}^k(\mathcal{M}_{/\mathbb{Z}_p})$$

is a cycle class which maps to  $\mathcal{C}_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{Q}_p$  under the homomorphism

$$\text{CH}_{\mathcal{Y}_{/\mathbb{Z}_p}}^k(\mathcal{M}_{/\mathbb{Z}_p}) \rightarrow Z_{\mathcal{Y}_{/\mathbb{Q}_p}}^k(\mathcal{M}_{/\mathbb{Q}_p}) \tag{12}$$

deduced from (7). The image of  $\mathcal{C}_p$  under

$$\mathrm{CH}_{\mathcal{Y}/\mathbb{Z}_p}^k(\mathcal{M}/\mathbb{Z}_p) \rightarrow \mathrm{CH}_{\mathrm{ver}}^k(\mathcal{M}/\mathbb{Z}_p),$$

denoted  $\mathcal{C}_p^{\mathrm{ver}}$ , is the *vertical component* of the local cycle datum. By a *global cycle datum* of codimension  $k$  with support on  $\mathcal{Y}$  we mean a triple  $(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_{\bullet})$  in which  $\mathcal{C}_{\mathbb{Q}}$  and  $\Xi$  are as above and  $\mathcal{C}_{\bullet} = \{\mathcal{C}_p\}$  is a collection indexed by the primes such that each  $(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_p)$  is a local cycle datum at  $p$  of codimension  $k$  with support along  $\mathcal{Y}$ . We further require that  $\mathcal{C}_p^{\mathrm{ver}} = 0$  for all but finitely many  $p$ . Any global cycle datum  $(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_{\bullet})$  determines an arithmetic cycle class

$$\widehat{\mathcal{C}}^{\mathrm{hor}} + \sum_p \widehat{\mathcal{C}}_p^{\mathrm{ver}} \in \widehat{\mathrm{CH}}^k(\mathcal{M}) \tag{13}$$

in which  $\widehat{\mathcal{C}}_p^{\mathrm{ver}}$  is the image of  $\mathcal{C}_p^{\mathrm{ver}}$  under the map (11) and  $\widehat{\mathcal{C}}^{\mathrm{hor}}$  is the arithmetic cycle consisting of the Zariski closure of  $\mathcal{C}_{\mathbb{Q}}$  in  $\mathcal{M}$  with its Green current  $\Xi$ .

Set  $\mathcal{Y}_0 = \mathcal{Y} \times_{\mathcal{M}} \mathcal{M}_0$  and assume that the image of  $\mathcal{Y}_0/\mathbb{Q} \rightarrow \mathcal{M}_0/\mathbb{Q}$  again has codimension  $k$ . There is a pullback global cycle datum  $(i^*\mathcal{C}_{\mathbb{Q}}, i^*\Xi, i^*\mathcal{C}_{\bullet})$  of codimension  $k$  with support along  $\mathcal{Y}_0$  defined as follows. For every prime  $p$  choose a  $\mathbb{Z}_p$ -uniformization  $\mathcal{M}/\mathbb{Z}_p \cong [H \backslash M]$  and define  $\mathbb{Z}_p$ -schemes

$$\begin{aligned} Y &= \mathcal{Y} \times_{\mathcal{M}} M, \\ M_0 &= \mathcal{M}_0 \times_{\mathcal{M}} M, \\ Y_0 &= \mathcal{Y}_0 \times_{\mathcal{M}_0} M_0. \end{aligned} \tag{14}$$

There is a pullback map

$$i^* : \mathrm{CH}_{\mathcal{Y}/\mathbb{Z}_p}^k(\mathcal{M}/\mathbb{Z}_p) \rightarrow \mathrm{CH}_{\mathcal{Y}_0/\mathbb{Z}_p}^k(\mathcal{M}_0/\mathbb{Z}_p)$$

defined as the composition

$$\mathrm{CH}_{\mathcal{Y}/\mathbb{Z}_p}^k(\mathcal{M}/\mathbb{Z}_p) \cong \mathrm{CH}_Y^k(M)^H \rightarrow \mathrm{CH}_{Y_0}^k(M_0)^H \cong \mathrm{CH}_{\mathcal{Y}_0/\mathbb{Z}_p}^k(\mathcal{M}_0/\mathbb{Z}_p)$$

in which the middle arrow is (10). In the exact same way one defines a pullback

$$i^* : Z_{\mathcal{Y}/\mathbb{Q}}^k(\mathcal{M}/\mathbb{Q}) \cong \mathrm{CH}_{\mathcal{Y}/\mathbb{Q}}^k(\mathcal{M}/\mathbb{Q}) \rightarrow \mathrm{CH}_{\mathcal{Y}_0/\mathbb{Q}}^k(\mathcal{M}_0/\mathbb{Q}) \cong Z_{\mathcal{Y}_0/\mathbb{Q}}^k(\mathcal{M}_0/\mathbb{Q})$$

using the hypotheses on the codimensions of  $\mathcal{Y}/\mathbb{Q}$  and  $\mathcal{Y}_0/\mathbb{Q}$  for the isomorphisms. This defines  $i^*\mathcal{C}_{\mathbb{Q}}$  and  $i^*\mathcal{C}_{\bullet}$ . The current  $i^*\Xi$  is defined as in [Sou92, §§ II.3.2 and II.3.3]: after replacing  $\Xi$  by an equivalent Green current we may assume that  $\Xi$  is a Green form of logarithmic type on  $M_{\mathbb{Q}}$ , and the pullback current  $i^*\Xi$  is then just the usual pullback of  $\Xi$  to  $M_{0\mathbb{Q}} = \mathcal{M}_0 \times_{\mathcal{M}} M_{\mathbb{Q}}$  in the sense of differential forms.

As before let  $\mathcal{M}_0$  be an arithmetic stack, equidimensional of dimension  $d_0$ , and suppose we have a closed immersion  $i : \mathcal{M}_0 \rightarrow \mathcal{M}$ . Let  $\mathcal{Y} \rightarrow \mathcal{M}$  be a proper morphism of algebraic stacks such that the image of  $\mathcal{Y}/\mathbb{Q} \rightarrow \mathcal{M}/\mathbb{Q}$  has codimension  $d_0$  and assume that  $\mathcal{Y}_0/\mathbb{Q} = \emptyset$ , where we again set  $\mathcal{Y}_0 = \mathcal{Y} \times_{\mathcal{M}} \mathcal{M}_0$ . In particular, the image of  $\mathcal{Y}_0/\mathbb{Q} \rightarrow \mathcal{M}_0/\mathbb{Q}$  has codimension  $d_0$ . If  $(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_{\bullet})$  is a global cycle datum on  $\mathcal{M}$  of codimension  $d_0$  supported on  $\mathcal{Y}$  then we have defined above a pullback global cycle datum  $(i^*\mathcal{C}_{\mathbb{Q}}, i^*\Xi, i^*\mathcal{C}_{\bullet})$  on  $\mathcal{M}_0$  of codimension  $d_0$  supported on  $\mathcal{Y}_0$ . Denoting by

$$\widehat{\mathcal{C}} \in \widehat{\mathrm{CH}}^{d_0}(\mathcal{M}), \quad \widehat{\mathcal{C}}_0 \in \widehat{\mathrm{CH}}^{d_0}(\mathcal{M}_0)$$

the arithmetic cycle classes corresponding to these cycle data, one may easily check that

$$\widehat{\mathrm{deg}}_{\mathcal{M}_0} \widehat{\mathcal{C}} = \widehat{\mathrm{deg}} \widehat{\mathcal{C}}_0.$$

In other words,

$$\widehat{\text{deg}}_{\mathcal{M}_0}(\mathcal{C}_{\mathbb{Q}}, \Xi, \mathcal{C}_{\bullet}) = \widehat{\text{deg}}(i^*\mathcal{C}_{\mathbb{Q}}, i^*\Xi, i^*\mathcal{C}_{\bullet}). \tag{15}$$

### 3. A Shimura surface and its special cycles

We take [Mum65, ch. 6] as our basic reference for abelian schemes. We will be dealing with integral models of the Shimura varieties associated to the algebraic groups  $G_0$  and  $G$  of the introduction. These have been dealt with thoroughly in the literature: for the Shimura curve associated to  $G_0$  we point out the articles of Buzzard [Buz97] and Boutot and Carayol [BC92], and for the Shimura surface associated to  $G$  we point out the volume [BBGGLMS79], as well as the articles of Milne [Mil79], Kottwitz [Kot92], Rapoport and Zink [RZ82, RZ96], Boutot and Zink [BZ], Kudla and Rapoport [KR99], and Hida [Hid04]. These Shimura surfaces are closely related to classical Hilbert modular surfaces, whose integral models are dealt with in work of Rapoport [Rap78], Deligne and Pappas [DP94], Pappas [Pap95], Stamm [Sta97], and Vollaard [Vol05]. General references on Hilbert modular varieties include the books of Goren [Gor02], Hida [Hid06], and van der Geer [van88]. The theory of Shimura varieties, from Deligne’s points of view, can be found in the works of Milne [Mil90, Mil05].

#### 3.1 Moduli problems

Let  $S$  be a scheme. By a *QM abelian fourfold* over  $S$  we mean a pair  $\mathbf{A} = (A, i)$  in which  $A$  is an abelian scheme over  $S$  of relative dimension four, and  $i : \mathcal{O}_B \rightarrow \text{End}(A)$  is a ring homomorphism taking  $1 \mapsto 1$  and satisfying the *Kottwitz condition* (see, for example, [Kot92, § 5], [Hid04, § 7.1], or [Vol05]), everywhere locally on  $S$ . This last condition means that for every  $s \in S$  there is an open affine neighborhood  $U$  over which  $\text{Lie}(A)$  is a free  $\mathcal{O}_U$ -module satisfying the equality of polynomials in  $\mathcal{O}_U[x]$ ,

$$\text{char}_{\mathcal{O}_U}(i(b); \text{Lie}(A)) = f_b(x) \cdot f_b(x)^\sigma$$

for every  $b \in \mathcal{O}_B$ . Here

$$f_b(x) = (x - b)(x - b^t) \in \mathcal{O}_F[x]$$

is the reduced characteristic polynomial of  $b$ . Define  $\text{End}(\mathbf{A})$  to be the  $\mathcal{O}_F$ -algebra of endomorphisms of  $A$  which commute with the action of  $\mathcal{O}_B$ , and set  $\text{End}^0(\mathbf{A}) = \text{End}(\mathbf{A}) \otimes_{\mathcal{O}_F} F$ . For a good theory of the moduli of QM abelian fourfolds we must introduce polarization data as well. By a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold we mean a pair  $(\mathbf{A}, \lambda)$  in which  $\mathbf{A} = (A, i)$  is a QM abelian fourfold and  $\lambda : A \rightarrow A^\vee$  is a polarization satisfying:

- (a)  $\lambda \circ i(b^*) = i(b)^\vee \circ \lambda$  for every  $b \in \mathcal{O}_B$ ;
- (b) the kernel of  $\lambda$  is  $A[\mathfrak{D}]$ .

Define  $\mathcal{M}$  to be the category, fibered in groupoids over the category of schemes, whose objects are  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfolds. A morphism from  $(\mathbf{A}', \lambda')$  to  $(\mathbf{A}, \lambda)$ , defined over schemes  $S'$  and  $S$ , respectively, in the category  $\mathcal{M}$  is a commutative diagram

$$\begin{array}{ccc} A' & \longrightarrow & A \\ \downarrow & & \downarrow \\ S' & \longrightarrow & S \end{array}$$

such that the induced map  $A' \rightarrow A \times_S S'$  is an isomorphism of abelian schemes over  $S$  respecting the action of  $\mathcal{O}_B$  and identifying the polarizations  $\lambda$  and  $\lambda'$ . The category  $\mathcal{M}$  is an arithmetic stack (in the sense of Definition 2.3.1) of dimension three with geometrically connected fibers. Furthermore,  $\mathcal{M}$  is smooth over  $\mathbb{Z}[1/D]$ , where  $D = \text{disc}(F) \cdot \text{disc}(B_0)$ . The existence of a smooth projective models for  $\mathcal{M}$  over  $\mathbb{Z}_p$  for primes  $p \nmid D$  is explained in [Kot92, Mil79], and in much greater detail in Chapters 6–7 of [Hid04]. When  $p \mid \text{disc}(B_0)$ , the regularity of the integral model follows from the Cerednik–Drinfeld uniformization described in [BC92] and [RZ96]; see § 4.1. The higher dimensional case needed here is dealt with in [BZ] and [RZ96]. Finally, for primes dividing  $\text{disc}(F)$  one proves the existence of regular projective models using the methods of Deligne and Pappas [DP94], who treat the case of classical Hilbert–Blumenthal moduli (e.g., the case  $B \cong M_2(F)$ , which we exclude). Deligne and Pappas use a different type of polarization data in the statement of their moduli problem, and do not impose the Kottwitz condition. The equivalence of moduli problems of the type defined above and those considered by Deligne and Pappas is explained in the work of Vollaard [Vol05].

By a *QM abelian surface* over a scheme  $S$  we mean a pair  $\mathbf{A}_0 = (A_0, i_0)$  in which  $A_0$  is an abelian scheme over  $S$  of relative dimension two and  $i_0 : \mathcal{O}_{B_0} \rightarrow \text{End}(A_0)$  is a ring homomorphism satisfying  $1 \mapsto 1$ . Again we require that  $\mathbf{A}_0$  satisfies the Kottwitz condition everywhere locally on  $S$ : for any  $b_0 \in \mathcal{O}_{B_0}$  the equality of polynomials in  $\mathcal{O}_U[x]$ ,

$$\text{char}_{\mathcal{O}_U}(i(b_0), \text{Lie}(A_0)) = (x - b_0)(x - b_0^*),$$

holds over any sufficiently small open affine neighborhood  $U$  of any point  $s \in S$ . Define  $\text{End}(\mathbf{A}_0)$  and  $\text{End}^0(\mathbf{A}_0)$  as above. By a *principally polarized* QM abelian surface we mean a pair  $(\mathbf{A}_0, \lambda_0)$  in which  $\mathbf{A}_0$  is a QM abelian surface and  $\lambda_0 : A_0 \rightarrow A_0^\vee$  is a polarization satisfying:

- (a)  $\lambda \circ i_0(b^*) = i_0(b)^\vee \circ \lambda_0$  for every  $b \in \mathcal{O}_{B_0}$ ;
- (b)  $\lambda_0$  is an isomorphism.

We remark that in fact every QM abelian surface admits a unique such  $\lambda_0$ . This follows by combining the argument used in [Buz97] to prove the claim when  $\text{disc}(B_0)$  is invertible on the base, with [BC92, Proposition III.3.3]. Let  $\mathcal{M}_0$  be the category, fibered in groupoids over the category of schemes, whose objects are principally polarized QM abelian surfaces  $(\mathbf{A}_0, \lambda_0)$  over schemes. Morphisms are defined exactly as in  $\mathcal{M}$ . The category  $\mathcal{M}_0$  is then an arithmetic stack of dimension two with geometrically connected fibers, and is smooth over  $\mathbb{Z}[\text{disc}(B_0)^{-1}]$ . See [BC92, Buz97] and the references therein for details.

Let  $\mathbf{A}_0 = (A_0, i_0)$  be a QM abelian surface over a scheme  $S$  with a polarization  $\lambda_0$  as above. The abelian fourfold  $A_0 \otimes \mathcal{O}_F \cong \text{Hom}_{\mathbb{Z}}(\mathfrak{D}^{-1}, A_0)$  (see [Con04, § 7] for background on Serre’s tensor construction) is then equipped with commuting actions of  $\mathcal{O}_{B_0}$  and  $\mathcal{O}_F$ , and hence with an action  $i : \mathcal{O}_B \rightarrow \text{End}(A_0 \otimes \mathcal{O}_F)$ . To be very concrete, we may define  $A_0 \otimes \mathcal{O}_F = A_0 \times A_0$  with  $\mathcal{O}_{B_0}$  acting diagonally and  $\mathcal{O}_F$  acting through the ring homomorphism  $\kappa : \mathcal{O}_F \rightarrow M_2(\mathbb{Z})$  determined by the  $\mathbb{Z}$ -basis  $\{\varpi_1, \varpi_2\}$  of  $\mathcal{O}_F$ . We obtain a QM abelian fourfold

$$\mathbf{A}_0 \otimes \mathcal{O}_F = (A_0 \otimes \mathcal{O}_F, i)$$

and an isomorphism of  $\mathcal{O}_F$ -algebras

$$\text{End}(\mathbf{A}_0 \otimes \mathcal{O}_F) \cong \text{End}(\mathbf{A}_0) \otimes_{\mathbb{Z}} \mathcal{O}_F. \tag{16}$$

The identification  $A_0 \otimes \mathcal{O}_F = A_0 \times A_0$  determines an isomorphism

$$(A_0 \otimes \mathcal{O}_F)^\vee \cong A_0^\vee \times A_0^\vee$$

in which the action of  $\mathcal{O}_{B_0}$  on the right-hand side is again diagonal, but the action of  $\mathcal{O}_F$  is through the *transpose* of the homomorphism  $\kappa$ . This transpose  ${}^t\kappa : \mathcal{O}_F \rightarrow M_2(\mathbb{Z})$  is the embedding determined by the action of  $\mathcal{O}_F$  on  $\mathfrak{D}^{-1}$  with respect to the dual basis  $\{\varpi_1^\vee, \varpi_2^\vee\}$  relative to the trace form on  $F$ , and allows us to identify

$$(A_0 \otimes \mathcal{O}_F)^\vee \cong A_0^\vee \times A_0^\vee \cong A_0^\vee \otimes \mathfrak{D}^{-1}.$$

If we denote by  $\lambda_0 \otimes \mathcal{O}_F$  the isogeny

$$A_0 \otimes \mathcal{O}_F \xrightarrow{\lambda_0 \otimes 1} A_0^\vee \otimes \mathfrak{D}^{-1} \cong (A_0 \otimes \mathcal{O}_F)^\vee,$$

then the pair  $(\mathbf{A}_0 \otimes \mathcal{O}_F, \lambda_0 \otimes \mathcal{O}_F)$  is a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold. We check that  $\lambda_0 \otimes \mathcal{O}_F$  is a polarization. Let  $\Delta$  be the matrix of the trace form on  $\mathcal{O}_F$  relative to the basis  $\{\varpi_1, \varpi_2\}$ . Identifying  $A_0 \otimes \mathcal{O}_F \cong A_0 \times A_0$  and  $(A_0 \otimes \mathcal{O}_F)^\vee \cong A_0^\vee \times A_0^\vee$  as above, the isogeny  $\lambda_0 \otimes \mathcal{O}_F$  is identified with

$$A_0 \times A_0 \xrightarrow{\lambda_0 \times \lambda_0} A_0^\vee \times A_0^\vee \xrightarrow{\Delta} A_0^\vee \times A_0^\vee. \tag{17}$$

As  $\Delta$  is positive definite there is some positive integer multiple  $m\Delta = {}^tE\Delta'E$  with  $\Delta', E \in M_2(\mathbb{Z})$  and  $\Delta'$  diagonal with diagonal entries  $d_1, d_2 > 0$ . The pullback of the polarization

$$A_0 \times A_0 \xrightarrow{d_1\lambda_0 \times d_2\lambda_0} A_0^\vee \times A_0^\vee$$

by the isogeny

$$A_0 \times A_0 \xrightarrow{E} A_0 \times A_0$$

is then  $m$  times the isogeny (17). As  $m$  times the isogeny (17) is a polarization, so is (17). We now have a functor

$$i : \mathcal{M}_0 \rightarrow \mathcal{M}$$

defined by

$$(\mathbf{A}_0, \lambda_0) \mapsto (\mathbf{A}_0, \lambda_0) \otimes \mathcal{O}_F = (\mathbf{A}_0 \otimes \mathcal{O}_F, \lambda_0 \otimes \mathcal{O}_F).$$

This functor induces a proper morphism of algebraic stacks. Combining the properness with the following lemma (which implies that the morphism  $i$  is injective on geometric points) shows that  $i : \mathcal{M}_0 \rightarrow \mathcal{M}$  is a closed immersion.

LEMMA 3.1.1. *For any principally polarized QM abelian surfaces  $(\mathbf{A}_0, \lambda_0)$  and  $(\mathbf{A}'_0, \lambda'_0)$  over a common base scheme  $S$ , the natural function*

$$\text{Isom}_{\mathcal{M}_0(S)}((\mathbf{A}_0, \lambda_0), (\mathbf{A}'_0, \lambda'_0)) \rightarrow \text{Isom}_{\mathcal{M}(S)}((\mathbf{A}_0, \lambda_0) \otimes \mathcal{O}_F, (\mathbf{A}'_0, \lambda'_0) \otimes \mathcal{O}_F)$$

is a bijection.

*Proof.* Abbreviate  $\mathbf{A} = \mathbf{A}_0 \otimes \mathcal{O}_F$  and  $\lambda = \lambda_0 \otimes \mathcal{O}_F$  and similarly for  $\mathbf{A}'_0$  and  $\lambda'_0$ . The isomorphism of  $\mathcal{O}_F$ -modules

$$\text{Hom}(\mathbf{A}_0, \mathbf{A}'_0) \otimes \mathcal{O}_F \cong \text{Hom}(\mathbf{A}, \mathbf{A}')$$

proves the injectivity of the function in question. For surjectivity, suppose we start with an isomorphism  $f : (\mathbf{A}, \lambda) \cong (\mathbf{A}', \lambda')$  in  $\mathcal{M}(S)$ . The condition that  $f$  identifies  $\lambda$  with  $\lambda'$  is equivalent



to the commutativity of the diagram

$$\begin{array}{ccccccc}
 \mathbf{A}_0 \times \mathbf{A}_0 & \longrightarrow & \mathbf{A} & \xrightarrow{f} & \mathbf{A}' & \longrightarrow & \mathbf{A}'_0 \times \mathbf{A}'_0 \\
 \Delta \circ \lambda_0 \downarrow & & \downarrow \lambda & & \downarrow \lambda' & & \downarrow \Delta \circ \lambda'_0 \\
 \mathbf{A}_0^\vee \times \mathbf{A}_0^\vee & \longleftarrow & \mathbf{A}^\vee & \xleftarrow{f^\vee} & \mathbf{A}'^\vee & \longleftarrow & \mathbf{A}'_0{}^\vee \times \mathbf{A}'_0{}^\vee
 \end{array}$$

in which all horizontal arrows are isomorphisms. The composition of arrows along the top row is given by some matrix

$$\Phi \in \text{Hom}(\mathbf{A}_0, \mathbf{A}'_0) \otimes M_2(\mathbb{Z})$$

which lies in the  $\mathbb{Z}$ -submodule  $\text{Hom}(\mathbf{A}_0, \mathbf{A}'_0) \otimes \kappa(\mathcal{O}_F)$ . The composition of arrows in the bottom row, from right to left, is then given by the transpose dual

$${}^t\Phi^* \in \text{Hom}(\mathbf{A}'_0, \mathbf{A}_0) \otimes M_2(\mathbb{Z})$$

where  $\Phi^*$  denotes the entry-by-entry dual of the matrix  $\Phi$ . Using the relation  $\Delta \cdot \kappa(x) = {}^t\kappa(x) \cdot \Delta$  for all  $x \in \mathcal{O}_F$ , the commutativity of the above diagram is equivalent to

$$\begin{pmatrix} \lambda_0 & \\ & \lambda_0 \end{pmatrix} = \Phi^* \cdot \begin{pmatrix} \lambda'_0 & \\ & \lambda'_0 \end{pmatrix} \cdot \Phi. \tag{18}$$

We now choose a  $\rho \in \text{GL}_2(\mathbb{Q})$  such that

$$\rho \kappa(F) \rho^{-1} = \left\{ \begin{pmatrix} a & bD \\ b & a \end{pmatrix} \mid a, b \in \mathbb{Q} \right\}$$

with  $D > 1$  a square free integer. Suppose first that  $D \equiv 1 \pmod{4}$ . Then there are  $f_1, f_2 \in \text{Hom}(\mathbf{A}_0, \mathbf{A}'_0)$  for which

$$\rho \Phi \rho^{-1} = \frac{1}{2} \begin{pmatrix} 2f_1 + f_2 & f_2 D \\ f_2 & 2f_1 + f_2 \end{pmatrix}.$$

Conjugating both sides of (18) by  $\rho$  and comparing the upper left entries on each side yields

$$\begin{aligned}
 4\lambda_0 &= (2f_1 + f_2)^\vee \circ \lambda'_0 \circ (2f_1 + f_2) + f_2^\vee \circ \lambda'_0 \circ f_2 D \\
 &= \text{deg}(2f_1 + f_2)\lambda_0 + \text{deg}(f_2)D\lambda_0
 \end{aligned}$$

where the second equality uses the fact that the  $\mathbb{Z}$ -module of symmetric homomorphisms  $\mathbf{A}_0 \rightarrow \mathbf{A}_0^\vee$  satisfying  $\lambda_0 \circ i_0(b^*) = i_0(b)^\vee \circ \lambda_0$  for all  $b \in \mathcal{O}_{B_0}$  is free of rank one and generated by  $\lambda_0$ , with the polarizations corresponding to positive multiples of  $\lambda_0$ . From this equality it is clear that  $f_2 = 0$ , and hence

$$\Phi = \begin{pmatrix} f_1 & \\ & f_1 \end{pmatrix}$$

as desired. The case  $D \equiv 2, 3 \pmod{4}$  is similar. □

**LEMMA 3.1.2.** *If  $\mathbf{A}_0$  is a QM abelian surface over a connected scheme  $S$  then  $\text{End}^0(\mathbf{A}_0)$  is either  $\mathbb{Q}$ , a quadratic imaginary field, or a definite quaternion algebra over  $\mathbb{Q}$ . If  $\mathbf{A}$  is a QM abelian fourfold over  $S$  then  $\text{End}^0(\mathbf{A})$  is either  $F$ , a totally imaginary quadratic extension of  $F$ , or a totally definite quaternion algebra over  $F$ .*

*Proof.* Fix a geometric point  $s \rightarrow S$ , and let  $\mathbf{A} = (A, i)$  be a QM abelian fourfold over  $S$ . By [Mum65, Corollary 6.2], the reduction map

$$\text{End}^0(\mathbf{A}) \rightarrow \text{End}^0(\mathbf{A}_{/k(s)})$$

is injective, and so it suffices to treat the case in which  $S = \text{Spec}(k)$  with  $k$  a field. One then further reduces to the case of  $k$  of finite transcendence degree over its prime subfield, and by considering the Néron model of  $A$  over a valuation ring inside of  $k$  and again invoking the injectivity of the reduction map on endomorphism algebras, one is reduced to the case of  $k$  finite over its prime subfield. If  $\text{char}(k) = 0$  then  $A(\mathbb{C})$  is isogenous to  $(B \otimes_{\mathbb{Q}} \mathbb{R})/\mathcal{O}_B$  for some complex structure on  $B \otimes_{\mathbb{Q}} \mathbb{R}$ , and from this one can show that  $\text{End}(\mathbf{A})$  is either  $F$  or a quadratic imaginary extension of  $F$ . If  $\text{char}(k) \neq 0$  then [Mil79, Proposition 5.2] shows that either  $A$  is isogenous to the fourth power of a supersingular elliptic curve, or  $A$  has no isogeny factor isomorphic to a supersingular elliptic curve. In the first case  $\text{End}^0(\mathbf{A})$  is isomorphic to the quaternion algebra  $\overline{B}$  over  $F$  satisfying

$$[B] + [\overline{B}] = [H \otimes_{\mathbb{Q}} F]$$

in the Brauer group of  $F$ , where  $H$  is the rational quaternion algebra of discriminant  $p$ . In the second case  $\text{End}^0(\mathbf{A})$  is a quadratic imaginary extension of  $F$ . Note that in the statement of [Mil79, Proposition 5.2] it is assumed that  $\text{char}(k)$  is prime to both  $\text{disc}(F)$  and to  $\text{disc}(B_0)$ . This hypothesis is not used in the proof until the sentence ‘It splits  $B$  because. . .’ and so has no bearing on the results cited above. The case of a QM abelian surface is identical (in this case one may also invoke [BC92, Proposition III.2] for primes dividing  $\text{disc}(B_0)$ ).  $\square$

Fix a finite set of rational primes  $\Sigma$  and let  $S$  be a scheme such that

$$\ell \notin \Sigma \implies \ell^{-1} \in \mathcal{O}_S.$$

Let  $U_0 \subset U_0^{\max}$  be a compact open subgroup which factors as  $U_0 = \prod_{\ell} U_{0,\ell}$  with  $U_{0,\ell} = U_0^{\max}$  for every  $\ell \in \Sigma$ . Write

$$U_0^{\max,\Sigma} = \prod_{\ell \notin \Sigma} U_{0,\ell}^{\max} \quad U_0^{\Sigma} = \prod_{\ell \notin \Sigma} U_{0,\ell}.$$

Let  $\Lambda_0 = \mathcal{O}_{B_0}$  be viewed as a left  $\mathcal{O}_{B_0}$ -module, and define a perfect alternating bilinear form  $\psi_0 : \Lambda_0 \times \Lambda_0 \rightarrow \mathbb{Z}$  as on [BC92, p. 130] by

$$\psi_0(x, y) = \frac{1}{\text{disc}(B_0)} \text{Tr}(xsy^*)$$

where  $\text{Tr}$  is the reduced trace on  $B_0$  and  $s \in \mathcal{O}_{B_0}$  is the element fixed in the introduction. Extend  $\psi_0$  to an alternating form

$$\psi_0 : \widehat{\Lambda}_0^{\Sigma} \times \widehat{\Lambda}_0^{\Sigma} \rightarrow \widehat{\mathbb{Z}}^{\Sigma} \tag{19}$$

on the restricted topological product

$$\widehat{\Lambda}_0^{\Sigma} = \prod_{\ell \notin \Sigma} (\Lambda_0 \otimes_{\mathbb{Z}} \mathbb{Z}_{\ell}).$$

We now define the notion of a  $U_0$  level structure on a principally polarized QM abelian surface  $(\mathbf{A}_0, \lambda_0)$  defined over  $S$ . Fix a geometric point  $s \rightarrow S$  and let

$$\text{Ta}^{\Sigma}(A_0) = \prod_{\ell \notin \Sigma} \text{Ta}_{\ell}(A_{0/k(s)})$$

be the prime-to- $\Sigma$  adelic Tate module of  $A_0$ , equipped with the action of  $(\widehat{\mathcal{O}}_{B_0})^{\Sigma}$  defined by  $i_0$  as well as the action of the étale fundamental group  $\pi_1(S, s)$ . A  $U_0$  level structure on  $\mathbf{A}_0$  is then a  $U_0$  equivalence class of  $(\widehat{\mathcal{O}}_{B_0})^{\Sigma}$ -module isomorphisms

$$\nu_0 : \widehat{\Lambda}_0^{\Sigma} \rightarrow \text{Ta}^{\Sigma}(A_0)$$

(here  $\nu_0$  and  $\nu'_0$  are  $U_0$  equivalent if there is a  $u \in U_0$  such that  $\nu_0(x) = \nu'_0(xu)$  for all  $x \in \widehat{\Lambda}_0^\Sigma$ ) which satisfy:

(a)  $\nu_0$  identifies the Weil pairing

$$\mathrm{Ta}^\Sigma(A_0) \times \mathrm{Ta}^\Sigma(A_0) \rightarrow \widehat{\mathbb{Z}}^\Sigma(1)$$

induced by  $\lambda_0$  with the pairing (19) up to a  $(\widehat{\mathbb{Z}}^\Sigma)^\times$ -multiple;

(b) for any  $\sigma \in \pi_1(S, s)$  there is a  $u \in U_0$  such that  $\sigma \cdot \nu_0(x) = \nu_0(xu)$ ; that is to say, the equivalence class of  $\nu_0$  is defined over  $S$ .

Set  $\Lambda = \Lambda_0 \otimes_{\mathbb{Z}} \mathcal{O}_F$ . For a compact open subgroup  $U \subset U^{\max}$  (factorizable and maximal at primes in  $\Sigma$ ) one defines the notion of  $U$  level structure on a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold  $(\mathbf{A}, \lambda)$  in the same way, replacing  $\widehat{\Lambda}_0^\Sigma$  by  $\widehat{\Lambda}^\Sigma = \widehat{\Lambda}_0^\Sigma \otimes_{\mathbb{Z}} \mathcal{O}_F$  and replacing  $\psi_0$  with the alternating form

$$\psi : \Lambda \times \Lambda \rightarrow \mathbb{Z}$$

defined by extending  $\psi_0$  to a  $\mathcal{O}_F$ -bilinear form  $\psi_0 : \Lambda \times \Lambda \rightarrow \mathcal{O}_F$  and setting  $\psi = \mathrm{Tr}_{F/\mathbb{Q}} \circ \psi_0$ .

If  $S$  is a scheme and  $(\mathbf{A}_0, \lambda_0)$  is an object of  $\mathcal{M}_0(S)$  then the polarization  $\lambda_0$  induces a Rosati involution  $\tau \mapsto \tau^\dagger$  on  $\mathrm{End}^0(\mathbf{A}_0)$ . The trace of an endomorphism  $\tau \in \mathrm{End}^0(\mathbf{A}_0)$  is defined to be  $\mathrm{Tr}(\tau) = \tau + \tau^\dagger$ , and we define a bilinear form on the trace zero elements of  $\mathrm{End}^0(\mathbf{A}_0)$  (with values in the commutative subalgebra of Rosati-fixed elements) by

$$[\tau_1, \tau_2] = -\mathrm{Tr}(\tau_1\tau_2).$$

The associated quadratic form is denoted by  $Q_0(\tau) = -\tau^2$ . Similar remarks and definitions hold for any object  $(\mathbf{A}, \lambda)$  of  $\mathcal{M}(S)$ , and on the trace zero elements of  $\mathrm{End}^0(\mathbf{A})$  we have the quadratic form  $Q(\tau) = -\tau^2$ .

For every  $\alpha \in \mathcal{O}_F$  define a category  $\mathcal{Y}(\alpha)$ , fibered in groupoids over the category of schemes, whose objects are triples  $(\mathbf{A}, \lambda, t_\alpha)$  in which  $(\mathbf{A}, \lambda)$  is a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold over a scheme and  $t_\alpha$  is a trace zero element of  $\mathrm{End}(\mathbf{A})$  satisfying  $Q(t_\alpha) = \alpha$ . There is an obvious functor

$$\phi : \mathcal{Y}(\alpha) \rightarrow \mathcal{M}$$

which forgets the endomorphism  $t_\alpha$ . Using the methods of [Hid04, chs 6-7] to prove the relative representability of  $\phi$ , one can show that the category  $\mathcal{Y}(\alpha)$  is an algebraic stack over  $\mathbb{Z}$ , finite over  $\mathcal{M}$ .

LEMMA 3.1.3. *The stack  $\mathcal{Y}(\alpha)_{/\mathbb{Q}}$  is étale over  $\mathrm{Spec}(\mathbb{Q})$ . In particular, the generic fiber of  $\mathcal{Y}(\alpha)$  is reduced of dimension zero.*

*Proof.* Choose a prime  $\ell$  which does not divide  $\mathrm{disc}(B_0)$  and with the property that

$$\mathcal{O}_F[\sqrt{-\alpha}] \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \cong \mathbb{Z}_\ell^4.$$

We abbreviate  $W = W(\mathbb{F}_\ell^{\mathrm{alg}})$  for the ring of Witt vectors of  $\mathbb{F}_\ell^{\mathrm{alg}}$  and  $M = \mathrm{Frac}(W)$  for the fraction field of  $W$ . Suppose  $z : \mathrm{Spec}(\mathbb{F}_\ell^{\mathrm{alg}}) \rightarrow \mathcal{Y}(\alpha)$  is a geometric point corresponding to a triple  $(\mathbf{A}, \lambda, t_\alpha)$  and let  $\mathfrak{g}$  denote the  $\ell$ -Barsotti–Tate group of  $\mathbf{A}$ . The completion  $R$  of the strictly Henselian local ring of  $\mathcal{Y}(\alpha)$  at  $z$  classifies deformations of  $(\mathbf{A}, \lambda, t_\alpha)$  to Artinian local  $W$ -algebras with residue field  $\mathbb{F}_\ell^{\mathrm{alg}}$ , and by the Serre–Tate theorem (and [BBGGLMS79, p. 51] or [Vol05] to see that the polarization  $\lambda$  deforms uniquely through Artinian thickenings of  $\mathbb{F}_\ell^{\mathrm{alg}}$ ) the  $W$ -algebra  $R$

is identified with the deformation space of the  $\ell$ -Barsotti–Tate group  $\mathfrak{g}$  with its action of

$$(\mathcal{O}_B \otimes_{\mathcal{O}_F} \mathcal{O}_F[\sqrt{-\alpha}]) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \cong M_2(\mathbb{Z}_\ell^4).$$

Using the usual idempotents on the right-hand side, we may decompose  $\mathfrak{g} \cong \mathfrak{g}_0 \times \mathfrak{g}_0$  in which  $\mathfrak{g}_0$  is an  $\ell$ -Barsotti–Tate group of height four and dimension two with an action of  $\mathbb{Z}_\ell^4$ , and then  $R$  classifies deformations of  $\mathfrak{g}_0$  with its action of  $\mathbb{Z}_\ell^4$ . The Kottwitz condition on  $\mathbf{A}$  implies that the usual idempotents in  $\mathbb{Z}_\ell^4$  induce a decomposition  $\mathfrak{g}_0 \cong (\mathbb{Q}_\ell/\mathbb{Z}_\ell)^2 \times \mu_{\ell^\infty}^2$ , and the theory of Serre–Tate coordinates as in [Gor02, Theorem 4.2] or [Hid04, Theorem 8.9] then implies that  $R \cong W$ .

If  $Y \rightarrow \mathcal{Y}(\alpha)_{/W}$  is a finite étale morphism with  $Y$  a  $W$ -scheme then, by the previous paragraph, the local ring of  $Y$  at any closed point is isomorphic to  $W$ . It follows that  $Y$  itself is simply a disjoint union of copies of  $\text{Spec}(W)$ , and hence that  $Y_{/M}$  is a disjoint union of copies of  $\text{Spec}(M)$ . We deduce that  $\mathcal{Y}(\alpha)_{/M}$  is étale over  $\text{Spec}(M)$ , and it follows that  $\mathcal{Y}(\alpha)_{/\mathbb{Q}}$  is étale over  $\text{Spec}(\mathbb{Q})$ .  $\square$

*Remark 3.1.4.* By Lemma 3.1.2, for any object  $(\mathbf{A}, \lambda)$  of  $\mathcal{M}$  the  $F$ -valued quadratic form  $Q$  is totally positive on the trace zero elements of  $\text{End}^0(\mathbf{A})$ . Thus, for any  $\alpha \in \mathcal{O}_F$ ,

$$\mathcal{Y}(\alpha) \neq \emptyset \implies \alpha \text{ totally positive.}$$

For any  $T \in \text{Sym}_2(\mathbb{Z})^\vee$ , let  $\mathcal{Z}(T)$  be the category, fibered in groupoids over schemes, whose objects are quadruples  $(\mathbf{A}_0, \lambda_0, s_1, s_2)$  in which  $(\mathbf{A}_0, \lambda_0)$  is an object of  $\mathcal{M}_0$  and  $s_1, s_2 \in \text{End}(\mathbf{A}_0)$  are trace zero endomorphisms of  $\mathbf{A}_0$  which satisfy

$$T = \frac{1}{2} \begin{pmatrix} [s_1, s_1] & [s_1, s_2] \\ [s_2, s_1] & [s_2, s_2] \end{pmatrix}. \tag{20}$$

We also define

$$\mathcal{Y}_0(\alpha) = \mathcal{Y}(\alpha) \times_{\mathcal{M}} \mathcal{M}_0$$

so that an object of  $\mathcal{Y}_0(\alpha)$  is a triple  $(\mathbf{A}_0, \lambda_0, t_\alpha)$  in which  $(\mathbf{A}_0, \lambda_0)$  is an object of  $\mathcal{M}_0$  and  $t_\alpha$  is a trace zero element of  $\text{End}(\mathbf{A}_0 \otimes \mathcal{O}_F)$  satisfying  $Q(t_\alpha) = \alpha$ . Fixing such a triple over a connected scheme  $S$ , (16) implies that  $t_\alpha$  has the form

$$t_\alpha = s_1\varpi_1 + s_2\varpi_2$$

for some trace zero  $s_1, s_2 \in \text{End}(\mathbf{A}_0)$ . The condition  $Q(t_\alpha) = \alpha$  is equivalent to the relation

$$\alpha = -s_1^2\varpi_1^2 - (s_1s_2 + s_2s_1)\varpi_1\varpi_2 - s_2^2\varpi_2^2.$$

Recalling (4), this last relation is equivalent to

$$\frac{1}{2} \begin{pmatrix} [s_1, s_1] & [s_1, s_2] \\ [s_2, s_1] & [s_2, s_2] \end{pmatrix} \in \Sigma(\alpha).$$

It follows easily that for each  $\alpha \in \mathcal{O}_F$  the rule  $(\mathbf{A}_0, \lambda_0, t_\alpha) \mapsto (\mathbf{A}_0, \lambda_0, s_1, s_2)$  identifies

$$\mathcal{Y}_0(\alpha) \cong \bigsqcup_{T \in \Sigma(\alpha)} \mathcal{Z}(T). \tag{21}$$

**LEMMA 3.1.5.** *Suppose  $\alpha \in \mathcal{O}_F$  is totally positive and let  $T$  be an element of  $\Sigma(\alpha)$ .*

- (a) *If  $\mathcal{Z}(T) \neq \emptyset$  then  $T$  is positive semi-definite.*
- (b) *If  $\mathcal{Z}(T)_{/\mathbb{Q}} \neq \emptyset$  then  $\det(T) = 0$ .*

(c) If  $F(\sqrt{-\alpha})$  is not a biquadratic extension of  $\mathbb{Q}$  then  $\det(T) \neq 0$ .

In particular, if  $\mathcal{Y}_0(\alpha)_{/\mathbb{Q}} \neq \emptyset$  then  $F(\sqrt{-\alpha})/\mathbb{Q}$  is biquadratic.

*Proof.* For part (a), if  $\mathcal{Z}(T) \neq \emptyset$  then there is some object  $(\mathbf{A}_0, \lambda_0, s_1, s_2)$  of  $\mathcal{Z}(T)$  defined over a connected scheme, and Lemma 3.1.2 implies that the quadratic form represented by (20) is positive definite. Part (b) follows from [KRY06, Theorem 3.6.1]. For part (c) suppose that  $\det(T) = 0$ . Then there is a one-dimensional  $\mathbb{Q}$ -vector space  $V_0$  equipped with a  $\mathbb{Q}$ -valued symmetric bilinear form  $[\cdot, \cdot]$  and vectors  $s_1, s_2 \in V_0$  for which the relation (20) holds. Let  $V = V_0 \otimes_{\mathbb{Q}} F$  and extend  $[\cdot, \cdot]$  to an  $F$ -valued symmetric bilinear form on  $V$ . The relation (20) then implies that the vector  $t_\alpha = s_1\varpi_1 + s_2\varpi_2$  satisfies  $[t_\alpha, t_\alpha] = 2\alpha$ . As  $V_0$  is one-dimensional we may write  $s_1 = n_1s, s_2 = n_2s$  as  $\mathbb{Q}$ -multiples of a common vector  $s \in V_0$ , and then

$$2\alpha = [t_\alpha, t_\alpha] = [s(n_1\varpi_1 + n_2\varpi_2), s(n_1\varpi_1 + n_2\varpi_2)] = (n_1\varpi_1 + n_2\varpi_2)^2 \cdot [s, s]$$

shows that  $\alpha \in \mathbb{Q}^\times \cdot (F^\times)^2$ . Hence  $F(\sqrt{-\alpha})/\mathbb{Q}$  is biquadratic.

The final claim follows by combining parts (b) and (c) with the decomposition (21). □

For the remainder of §3 we fix a totally positive  $\alpha \in \mathcal{O}_F$  and abbreviate  $\mathcal{Y} = \mathcal{Y}(\alpha)$  and  $\mathcal{Y}_0 = \mathcal{Y}_0(\alpha)$  for the algebraic stacks over  $\mathbb{Z}$  constructed above. Define  $j$  and  $\phi_0$  by the requirement that the diagram

$$\begin{array}{ccc} \mathcal{Y}_0 & \xrightarrow{\phi_0} & \mathcal{M}_0 \\ j \downarrow & & \downarrow i \\ \mathcal{Y} & \xrightarrow{\phi} & \mathcal{M} \end{array}$$

is cartesian.

### 3.2 Green currents

Let  $\mathcal{M}_{/\mathbb{Q}} \cong [H \backslash M]$  be a  $\mathbb{Q}$ -uniformization of  $\mathcal{M}$ , define  $\mathbb{Q}$ -schemes

$$Y = \mathcal{Y} \times_{\mathcal{M}} M, \quad M_0 = \mathcal{M}_0 \times_{\mathcal{M}} M,$$

and let  $\phi : Y \rightarrow M$  and  $i : M_0 \rightarrow M$  be the projections to the second factors. By Lemma 3.1.3 the scheme  $Y$  is a disjoint union of reduced zero-dimensional components, and we define

$$C_{\mathbb{Q}} = \sum_{y \in Y} \phi(y) \in Z_Y^2(M). \tag{22}$$

As  $C_{\mathbb{Q}}$  is  $H$ -invariant it may be identified, using [Gil84, Lemma 4.3], with a cycle  $\mathcal{C}_{\mathbb{Q}} \in Z_{\mathcal{Y}_{/\mathbb{Q}}}^2(\mathcal{M}_{/\mathbb{Q}})$  which is independent of the choice of  $\mathbb{Q}$ -uniformization of  $\mathcal{M}$ .

The remainder of §3.2 is devoted to the construction of a Green current for  $\mathcal{C}_{\mathbb{Q}}$ . In order to construct such a current we will use the  $\mathbb{Q}$ -uniformizations of  $\mathcal{M}$  which come from the theory of canonical models of Shimura varieties. Let  $\text{Nm} : B^\times \rightarrow F^\times$  denote the reduced norm on  $B$ . Define a real algebraic group  $\mathbb{S} = \text{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m$  and pick a  $z \in G_0(\mathbb{R})$  with  $z^2 = -1$ . The point  $z$  determines a map of  $\mathbb{R}$ -algebras  $\mathbb{C} \rightarrow B_0 \otimes_{\mathbb{Q}} \mathbb{R}$  by  $a + bi \mapsto a + bz$ , which in turn determines a map of real algebraic groups  $h_z : \mathbb{S} \rightarrow G_{0/\mathbb{R}}$ . Let  $X_0$  denote the  $G_0(\mathbb{R})$ -conjugacy class of any such map (all are conjugate by the Noether–Skolem theorem). We may also view  $h_z$  as a map  $\mathbb{S} \rightarrow G_{/\mathbb{R}}$  and let  $X$  be the  $G(\mathbb{R})$ -conjugacy class of  $h_z$ . The set  $X_0$  is naturally identified with the subset of  $X$  consisting of those maps  $\mathbb{S} \rightarrow G_{/\mathbb{R}}$  which factor through the subgroup  $G_{0/\mathbb{R}}$ . The

isomorphism of  $\mathbb{R}$ -algebras (3) determines an isomorphism

$$B \otimes_{\mathbb{Q}} \mathbb{R} \cong (B_0 \otimes_{\mathbb{Q}} \mathbb{R}) \times (B_0 \otimes_{\mathbb{Q}} \mathbb{R}) \tag{23}$$

which in turn determines an injection  $G(\mathbb{R}) \rightarrow G_0(\mathbb{R}) \times G_0(\mathbb{R})$  and an injection  $X \rightarrow X_0 \times X_0$  taking the subset  $X_0 \subset X$  bijectively onto the diagonal. If we fix an isomorphism  $B_0 \otimes_{\mathbb{Q}} \mathbb{R} \cong M_2(\mathbb{R})$  then we may identify the set  $X_0$  with the complex manifold  $\mathbb{C} \setminus \mathbb{R}$  in the usual way [Mil05, Example 5.6]. The resulting complex manifold structure on  $X_0$  is independent of the choice of isomorphism  $B_0 \otimes_{\mathbb{Q}} \mathbb{R} \cong M_2(\mathbb{R})$ . The topological space  $X_0$  has two connected components, say  $X_0 = X_0^+ \cup X_0^-$ , and the injection  $X \rightarrow X_0 \times X_0$  identifies

$$X \cong (X_0^+ \times X_0^+) \cup (X_0^- \times X_0^-). \tag{24}$$

The groups  $G_0(\mathbb{R})$  and  $G(\mathbb{R})$  act on  $X_0$  and  $X$ , respectively, through conjugation. In particular, the complex manifold  $X \times G(\mathbb{A}_f)$  admits a left action by  $G(\mathbb{Q})$  (acting on both factors) and a right action (on the second factor) by the maximal compact open subgroup  $U^{\max} \subset G(\mathbb{A}_f)$ .

Fix a normal compact open subgroup  $U \subset U^{\max}$  small enough that the algebraic stack of  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfolds over  $\mathbb{Q}$  with level  $U$  structure is a scheme,  $M$ , so that there is a  $\mathbb{Q}$ -uniformization of  $\mathcal{M}$ ,

$$\mathcal{M}/_{\mathbb{Q}} \cong [H \backslash M]$$

where  $H = U^{\max}/U$ . By Shimura’s theory [Mil79, Mil05] there is an isomorphism of complex manifolds

$$M(\mathbb{C}) \cong G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f) / U. \tag{25}$$

To (briefly) make the isomorphism (25) explicit, fix a point  $(x, g) \in X \times G(\mathbb{A}_f)$  and use strong approximation to factor  $g = \gamma u$  with  $\gamma \in G(\mathbb{Q})$  and  $u \in U^{\max}$ . The point  $\gamma^{-1}x \in X$  determines (in fact, is) a map of real algebraic groups  $\mathbb{S} \rightarrow G/\mathbb{R}$ . Setting  $\Lambda = \mathcal{O}_B$ , the group  $G(\mathbb{R})$  acts by right multiplication on  $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ , and the induced Hodge structure

$$\mathbb{S} \rightarrow \text{Aut}_{\mathbb{R}}(\Lambda \otimes_{\mathbb{Z}} \mathbb{R})$$

determines a complex structure on the real vector space  $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ . Recall the alternating pairing  $\psi : \Lambda \times \Lambda \rightarrow \mathbb{Z}$  of § 3.1. It follows from [Mil79, Lemma 1.1] that either  $\psi$  or  $-\psi$ , depending on the connected component of  $X$  containing  $\gamma^{-1}x$ , determines a  $\mathfrak{D}^{-1}$ -polarization  $\lambda$  of the complex torus  $A = (\Lambda \otimes_{\mathbb{Z}} \mathbb{R})/\Lambda$  with its left  $\mathcal{O}_B$ -action, and, in the notation of § 3.1, the isomorphism of left  $\widehat{\mathcal{O}}_B$ -modules

$$\nu : \widehat{\Lambda}^{\theta} \xrightarrow{u} \widehat{\Lambda}^{\theta} \cong \text{Ta}^{\theta}(A)$$

is a  $U$  level structure on  $\mathbf{A} = (A, i)$ . The isomorphism (25) then identifies the double coset of  $(x, g)$  with the isomorphism class of the triple  $(\mathbf{A}, \lambda, \nu)$ . Similarly, one can show that

$$M_0(\mathbb{C}) \cong G_0(\mathbb{Q}) \backslash X_0 \times G_0(\mathbb{A}_f) U^{\max} / U.$$

Let  $V_0$  and  $V$  denote the trace zero elements of  $B_0$  and  $B$ , respectively, with  $G_0(\mathbb{Q})$  and  $G(\mathbb{Q})$  acting on  $V_0$  and  $V$  by conjugation. Then  $V$  is endowed with the  $G(\mathbb{Q})$ -invariant  $F$ -valued quadratic form  $Q(\tau) = -\tau^2$ , and  $V_0$  is endowed with the  $G_0(\mathbb{Q})$ -invariant  $\mathbb{Q}$ -valued quadratic form  $Q_0$  defined by the same formula. Each nonzero  $\tau \in V_0 \otimes_{\mathbb{Q}} \mathbb{R}$ , viewed as an element of  $G_0(\mathbb{R})$ , acts by conjugation on  $X_0$  with fixed point set

$$X_0(\tau) = \{x \in X_0 \mid \tau \cdot x = x\}$$

consisting of two points if  $Q_0(\tau)$  is positive and no points otherwise. Assuming that  $Q_0(\tau)$  is positive, let  $x_0^+(\tau) \in X_0^+$  and  $x_0^-(\tau) \in X_0^-$  be the two points in  $X_0(\tau)$ . Identify  $V \otimes_{\mathbb{Q}} \mathbb{R} \cong$

$(V_0 \otimes_{\mathbb{Q}} \mathbb{R})^2$  using (23). If  $\tau = (\tau_1, \tau_2) \in V \otimes_{\mathbb{Q}} \mathbb{R}$  with  $Q(\tau)$  totally positive, then we define points  $x^+(\tau), x^-(\tau) \in X$  by

$$x^\pm(\tau) = (x_0^\pm(\tau_1), x_0^\pm(\tau_2))$$

using the isomorphism of (24), and set  $X(\tau) = \{x^+(\tau), x^-(\tau)\}$ . If  $Q(\tau)$  is not totally positive then set  $X(\tau) = \emptyset$ .

Given a positive parameter  $u \in \mathbb{R}$ , Kudla has defined (see [KRY06, Proposition 7.3.1 and (7.3.42)]; these functions were first constructed in [Kud97]) a symmetric  $G_0(\mathbb{R})$ -invariant Green function  $g_u^0(z_1, z_2)$  for the diagonal  $X_0 \subset X_0 \times X_0$ , identically zero off of the subset  $X \subset X_0 \times X_0$ . For each  $\tau_0 \in V_0 \otimes_{\mathbb{Q}} \mathbb{R}$  with  $Q_0(\tau_0) > 0$ , define a function  $\xi_0(\tau_0)$  on  $X_0$  by

$$\xi_0(\tau_0)(z) = \sum_{x \in X_0(\tau_0)} g_{Q_0(\tau_0)}^0(x, z).$$

The function  $\xi_0(\tau_0)$  is a Green function for the 0-cycle  $X_0(\tau_0)$  on  $X_0$ . Now suppose  $\tau \in V \otimes_{\mathbb{Q}} \mathbb{R}$  with  $Q(\tau)$  totally positive and write  $(\tau_1, \tau_2)$  for the corresponding element of  $(V_0 \otimes_{\mathbb{Q}} \mathbb{R})^2$ . Define functions  $\xi^1(\tau), \xi^2(\tau)$  in the variables  $(z_1, z_2) \in X_0 \times X_0$  by

$$\xi^1(\tau)(z_1, z_2) = \xi_0(\tau_1)(z_1), \quad \xi^2(\tau)(z_1, z_2) = \xi_0(\tau_2)(z_2).$$

These functions are Green functions for the divisors  $X_0(\tau_1) \times X_0$  and  $X_0 \times X_0(\tau_2)$  (respectively) on  $X_0 \times X_0$ . It follows that the star product [Sou92, II.3] of Green functions on  $X_0 \times X_0$ ,

$$\xi(\tau) = \xi^1(\tau) * \xi^2(\tau),$$

is a Green current for the 0-cycle  $X_0(\tau_1) \times X_0(\tau_2) \subset X_0 \times X_0$ , and restricting to a current on  $X$  yields a Green current for the 0-cycle  $X(\tau)$ .

Now fix a totally positive  $v \in F \otimes_{\mathbb{Q}} \mathbb{R}$  and write  $v^{1/2}$  for the totally positive square root of  $v$ . Define an  $\mathcal{O}_F$ -lattice in  $V$  by  $L = V \cap \mathcal{O}_B$  and a  $(1, 1)$ -current on  $X$ ,

$$\Xi_v(\alpha) = \sum_{\substack{\tau \in L \\ Q(\tau) = \alpha}} \xi(v^{1/2}\tau).$$

For simplicity, we abbreviate  $\Xi_v = \Xi_v(\alpha)$ . It follows from the  $G_0(\mathbb{R})$ -invariance of the Green function  $g_u^0(z_1, z_2)$  that

$$\gamma^* \xi(\tau) = \xi(\gamma \cdot \tau)$$

for any  $\gamma \in G(\mathbb{R})$  and  $\tau \in V$ . From this and the stability of  $L$  under  $\Gamma^{\max}$  one sees that  $\Xi_v$  is  $\Gamma^{\max}$ -invariant. Viewing  $\Xi_v$  as a current on the subset  $X \times \{1\} \subset X \times G(\mathbb{A}_f)$  and using the factorization  $G(\mathbb{A}_f) = G(\mathbb{Q})U^{\max}$  of strong approximation, we see that  $\Xi_v$  extends uniquely to a left  $G(\mathbb{Q})$ -invariant and right  $U^{\max}$ -invariant current on  $X \times G(\mathbb{A}_f)$ , which then descends to an  $H$ -invariant current on  $M(\mathbb{C})$ . If for each  $\tau \in V$  we define

$$\Omega(\tau) = \{g \in G(\mathbb{A}_f) \mid \tau \in g\widehat{\mathcal{O}}_B g^{-1}\}$$

then there is a bijection of zero-dimensional complex manifolds

$$Y(\mathbb{C}) \cong G(\mathbb{Q}) \backslash \bigsqcup_{\substack{\tau \in V \\ Q(\tau) = \alpha}} (X(\tau) \times \Omega(\tau)) / U \tag{26}$$

for which the map  $\phi : Y \rightarrow M$  is induced by the evident map

$$X(\tau) \times \Omega(\tau) \rightarrow X \times G(\mathbb{A}_f).$$

Here the action of  $G(\mathbb{Q})$  permutes the summands in the disjoint union through the conjugation action of  $G(\mathbb{Q})$  on  $V$ . Indeed, suppose we are given a triple  $(\tau, x, g)$  with  $\tau \in V$ ,  $Q(\tau) = \alpha$ , and  $(x, g) \in X(\tau) \times \Omega(\tau)$ . We again factor  $g = \gamma u$  with  $\gamma \in G(\mathbb{Q})$  and  $u \in U^{\max}$ , so that  $(\tau, x, g)$  and  $(\gamma^{-1}\tau, \gamma^{-1}x, u)$  have the same image in the right-hand side of (26). To  $(\tau, x, g)$  we associate the quadruple  $(\mathbf{A}, \lambda, \nu, t_\alpha)$  in which  $(\mathbf{A}, \lambda, \nu)$  is the  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold over  $\mathbb{C}$  with  $U$  level structure attached to  $(x, g)$  as above, and  $t_\alpha$  is the trace zero endomorphism of  $\mathbf{A}$ ,

$$t_\alpha : (\Lambda \otimes_{\mathbb{Z}} \mathbb{R})/\Lambda \rightarrow (\Lambda \otimes_{\mathbb{Z}} \mathbb{R})/\Lambda,$$

defined by  $t_\alpha(b) = b \cdot (\gamma^{-1}\tau)$ . Note that the  $\cdot$  is multiplication in  $B$ .

If we define an infinite formal sum on  $X$ ,

$$\mathcal{D} = \sum_{\substack{\tau \in L \\ Q(\tau) = \alpha}} \sum_{x \in X(\tau)} x,$$

then  $\mathcal{D}$ , viewed as a formal sum on  $X \times \{1\}$ , extends uniquely to a left  $G(\mathbb{Q})$ -invariant and right  $U^{\max}$ -invariant formal sum on  $X \times G(\mathbb{A}_f)$  whose associated 0-cycle on  $M(\mathbb{C})$  is (22). Comparing  $\mathcal{D}$  with the definition of  $\Xi_v$  and using the fact that  $\xi(v^{1/2}\tau)$  is a Green current for the 0-cycle  $X(\tau)$  shows that  $\Xi_v$  is a Green current for  $\mathcal{C}_{\mathbb{Q}} \in Z^2(\mathcal{M}/\mathbb{Q})$ .

The quadratic form  $Q_0$  determines a bilinear form  $[s_1, s_2] = -\text{Tr}(s_1 s_2)$  on  $V_0$ , where  $\text{Tr}$  is the reduced trace on  $B_0$ . For any  $T \in \Sigma(\alpha)$ , let  $V_0(T) \subset V_0 \times V_0$  be the set of pairs  $(s_1, s_2)$  for which (20) holds. Given any  $(s_1, s_2) \in V_0(T)$ , set

$$\tau = s_1 \varpi_1 + s_2 \varpi_2 \in V$$

and let  $\tau_1, \tau_2 \in V_0 \otimes_{\mathbb{Q}} \mathbb{R}$  be defined as above. Define  $L_0 = V_0 \cap \mathcal{O}_{B_0}$  and  $L_0(T) = (L_0 \times L_0) \cap V_0(T)$ . Let  $v = (v_1, v_2)$  under the isomorphism (3) and let  $T \in \Sigma(\alpha)$  be nonsingular. Then  $\{s_1, s_2\}$  is a linearly independent set, as is  $\{\tau_1, \tau_2\}$  which implies that the 0-cycles  $X_0(\tau_1)$  and  $X_0(\tau_2)$  on  $X_0$  have no common components. Hence the star product  $\xi_0(v_1^{1/2} \tau_1) * \xi_0(v_2^{1/2} \tau_2)$  is a Green current for the empty cycle on  $X_0$ .

**PROPOSITION 3.2.1.** *Assume that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic, so that, by Lemma 3.1.5,  $\Sigma(\alpha)$  contains no singular matrices and  $\mathcal{Y}_{0/\mathbb{Q}} = \emptyset$ . Then*

$$\text{deg}_{\infty}(i^* \Xi_v) = \frac{1}{2} \sum_{\substack{T \in \Sigma(\alpha) \\ (s_1, s_2) \in \Gamma_0^{\max} \setminus L_0(T)}} \frac{1}{|\text{stab}(s_1, s_2)|} \int_{X_0} \xi_0(v_1^{1/2} \tau_1) * \xi_0(v_2^{1/2} \tau_2)$$

where  $\text{stab}(s_1, s_2)$  is the stabilizer of  $(s_1, s_2)$  in  $\Gamma_0^{\max}$ .



*Proof.* Formally one has

$$\begin{aligned}
 \deg_\infty(i^*\Xi_v) &= \frac{1}{2 \cdot [\Gamma_0^{\max} : \Gamma_0]} \int_{\Gamma_0 \backslash X_0} i^*\Xi_v \\
 &= \frac{1}{2 \cdot [\Gamma_0^{\max} : \Gamma_0]} \sum_{\substack{\tau \in \Gamma_0 \backslash L \\ Q(\tau)=\alpha}} \int_{X_0} i^*(\xi^1(v^{1/2}\tau) * \xi^2(v^{1/2}\tau)) \\
 &= \frac{1}{2 \cdot [\Gamma_0^{\max} : \Gamma_0]} \sum_{\substack{\tau \in \Gamma_0 \backslash L \\ Q(\tau)=\alpha}} \int_{X_0} (i^*\xi^1(v^{1/2}\tau)) * (i^*\xi^2(v^{1/2}\tau)) \\
 &= \frac{1}{2 \cdot [\Gamma_0^{\max} : \Gamma_0]} \sum_{\substack{\tau \in \Gamma_0 \backslash L \\ Q(\tau)=\alpha}} \int_{X_0} \xi_0(v_1^{1/2}\tau_1) * \xi_0(v_2^{1/2}\tau_2) \tag{27}
 \end{aligned}$$

where  $\Gamma_0 = \Gamma \cap \Gamma_0^{\max}$  for any sufficiently small subgroup  $\Gamma \subset \Gamma^{\max}$  of finite index. These calculations require some caution, as the definition of pullback of currents is slightly subtle. Recall from § 2.3 that to define  $i^*\Xi_v$  we first replace the Green current  $\Xi_v$  on  $\Gamma \backslash X$  by an equivalent Green form of logarithmic type along (the pullback to  $\Gamma \backslash X$  of) the 0-cycle  $\mathcal{Y}(\alpha)(\mathbb{C})$ . Such a form exists by [Sou92, II.3.3], but the argument given there makes use of the compactness of  $\Gamma_0 \backslash X_0$ . As we cannot apply this argument to currents on  $X$ , we in fact do not even have a definition of the current  $i^*(\xi^1(v_1^{1/2}\tau_1) * \xi^2(v_2^{1/2}\tau_2))$  on  $X_0$ . We avoid this issue by making use of the methods of [GS91, § 2.1.5]. For each  $\epsilon > 0$  one can construct from Kudla’s Green function  $g_u^0$  a smooth  $\Gamma_0$ -invariant  $(1, 1)$ -form  $\omega_\epsilon$  on  $X$  which is supported on the hyperbolic tube of radius  $\epsilon$  around  $X_0$ , satisfies  $\lim_{\epsilon \rightarrow 0} [\omega_\epsilon] = \delta_{X_0}$  as currents on  $X$ , and such that for each  $P \in X_0$  the restriction of  $\omega_\epsilon$  to  $X_0 \times \{P\}$  (as well as to  $\{P\} \times X_0$ ) converges to the current  $\delta_P$  on  $X_0$ . The integral in the second line of (27) is then interpreted as the limit as  $\epsilon \rightarrow 0$  of the current  $\xi^1(v_1^{1/2}\tau_1) * \xi^2(v_2^{1/2}\tau_2)$  evaluated at the smooth form  $\omega_\epsilon$ . If we now define a smooth  $(1, 1)$ -form

$$\Omega_\epsilon = \sum_{\gamma \in \Gamma_0 \backslash \Gamma} \gamma^* \omega_\epsilon$$

on  $\Gamma \backslash X$  then  $\lim_{\epsilon \rightarrow 0} [\Omega_\epsilon] = \delta_{\Gamma_0 \backslash X_0}$  as currents on  $\Gamma \backslash X$ , and the integral in the first line of (27) is equal to the limit as  $\epsilon \rightarrow 0$  of the value of the current  $\Xi_v$  on the form  $\Omega_\epsilon$ . From this the second equality of (27) is clear. It is now an easy exercise in the definition of the star product to verify that if we are given Green functions of logarithmic type  $a_0$  and  $b_0$  for disjoint divisors on  $X_0$  and define  $a = \pi_1^* a_0$  and  $b = \pi_2^* b_0$  (in which  $\pi_1, \pi_2 : X \rightarrow X_0$  are the two projections), then

$$\int_X i^*(a * b) = \int_X (i^*a) * (i^*b) = \int_{X_0} a_0 * b_0$$

(compare with [GS91, p. 159]). Taking

$$a_0 = \xi_0(v_1^{1/2}\tau_1), \quad b_0 = \xi_0(v_2^{1/2}\tau_2)$$

yields the final two equalities of (27).

We have now established

$$\deg_\infty(i^*\Xi_v) = \frac{1}{2 \cdot [\Gamma_0^{\max} : \Gamma_0]} \sum_{\substack{\tau \in \Gamma_0 \backslash L \\ Q(\tau)=\alpha}} \int_{X_0} \xi_0(v_1^{1/2}\tau_1) * \xi_0(v_2^{1/2}\tau_2).$$

The  $\Gamma_0$ -invariant function

$$L_0 \times L_0 \rightarrow L_0 \otimes_{\mathbb{Z}} \mathcal{O}_F \cong L$$

defined by  $(s_1, s_2) \mapsto s_1\varpi_1 + s_2\varpi_2$  restricts to a  $\Gamma_0$ -invariant bijection

$$\bigcup_{T \in \Sigma(\alpha)} L_0(T) \rightarrow \{\tau \in L \mid Q(\tau) = \alpha\},$$

and the proposition follows easily. □

### 3.3 Cycle classes at primes of good reduction

Let  $p$  be a prime at which  $B_0$  is unramified. Fix a  $\mathbb{Z}_p$ -uniformization  $\mathcal{M}/\mathbb{Z}_p \cong [H \backslash M]$  of  $\mathcal{M}$  and define  $\mathbb{Z}_p$ -schemes  $Y$ ,  $M_0$ , and  $Y_0$  by (14). Let  $\phi : Y \rightarrow M$  be the projection.

PROPOSITION 3.3.1. *The scheme  $Y$  is at most one-dimensional. In particular,*

$$\mathbf{K}_0^Y(M) = F^2 \mathbf{K}_0^Y(M).$$

*Proof.* The horizontal components of  $Y$  are all Zariski closures of points of the (zero-dimensional) generic fiber  $Y/\mathbb{Q}_p$ , and so are of dimension one. We must therefore show that  $Y/\mathbb{F}_p^{\text{alg}}$  has dimension at most one. One can easily show that the Serre–Tate canonical lift gives an injection from the set of ordinary points of  $Y(\mathbb{F}_p^{\text{alg}})$  to the finite set  $Y(W(\mathbb{F}_p^{\text{alg}}))$ , and so it suffices to show that the locus of nonordinary points of  $Y/\mathbb{F}_p^{\text{alg}}$  is of dimension at most one. As the map  $Y \rightarrow M$  is finite, we are now reduced to proving that the nonordinary locus of  $M/\mathbb{F}_p^{\text{alg}}$  is of dimension one. We give only a sketch of the proof; the idea is to reduce the question to the setting of the Hilbert–Blumenthal surface where it follows from calculations of Goren and others [AG03, AG04, AG05, GO00]. Briefly, an abelian variety  $A$  over  $\mathbb{F}_p^{\text{alg}}$  is ordinary if and only if the Verschiebung  $\text{Ver} : \text{Lie}(A^{(p)}) \rightarrow \text{Lie}(A)$  is an isomorphism of  $\mathbb{F}_p^{\text{alg}}$ -modules. Thus the nonordinary locus of  $M/\mathbb{F}_p^{\text{alg}}$  is defined locally by a single equation, the determinant of the Hasse–Witt matrix encoding the action of Verschiebung on Lie algebras.

Fix a nonordinary point  $x \in M(\mathbb{F}_p^{\text{alg}})$  and denote by  $(\mathbf{A}, \lambda)$  the  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold determined by  $x$ . Let  $R$  be the completion of the local ring of  $M/\mathbb{F}_p^{\text{alg}}$  at  $x$ . By the Serre–Tate theory the functor of deformations of the QM-abelian surface  $(\mathbf{A}, \lambda)$  to local Artinian  $\mathbb{F}_p^{\text{alg}}$ -algebras with residue field  $\mathbb{F}_p^{\text{alg}}$  is isomorphic to the functor of deformations of the associated polarized  $p$ -Barsotti–Tate group  $(\mathbf{A}_p, \lambda_p)$  of height eight and dimension four with its action of  $\mathcal{O}_B \otimes_{\mathcal{O}_F} \mathcal{O}_{F,p} \cong M_2(\mathcal{O}_{F,p})$ , and this functor is pro-represented by  $R$ . Here we have set  $\mathcal{O}_{F,p} = \mathcal{O}_F \otimes_{\mathbb{Z}} \mathbb{Z}_p$ . Using the action of the idempotents in  $M_2(\mathcal{O}_{F,p})$  to decompose  $\mathbf{A}_p$  one can identify this deformation functor with the deformation functor of a quasi-polarized  $p$ -Barsotti–Tate group  $(\mathbf{B}_p, \eta_p)$  of height four and dimension two equipped with an action of  $\mathcal{O}_{F,p}$ . This is exactly the type of deformation problem considered by Goren *et al.* in their study of local rings of Hilbert–Blumenthal surfaces.

When  $p$  is inert in  $F$  then contemplation of the possible slope sequences shows that  $\mathbf{B}_p$  is supersingular. One can apply Zink’s theory of displays [Zin02] of connected  $p$ -Barsotti–Tate groups exactly as in [Gor02, ch. 6] or [GO00] to show that  $R$  is a power series ring in two variables over  $\mathbb{F}_p^{\text{alg}}$ , while the determinant  $r \in R$  of the Hasse–Witt matrix of the universal deformation is nonzero. The quotient ring  $R/(r)$  is therefore of dimension one as desired. When  $p$  is ramified in  $F$  then again  $\mathbf{B}_p$  is supersingular. In this case the  $\mathbb{F}_p^{\text{alg}}$ -algebra  $R$  may not be smooth, but one can deduce from calculations of [AG03] that the quotient of  $R$  by the determinant of the Hasse–

Witt matrix has dimension one. When  $p$  splits in  $F$  then the idempotents in  $\mathcal{O}_{F,p} \cong \mathbb{Z}_p \times \mathbb{Z}_p$  further split  $(\mathbf{B}_p, \eta_p)$  into a product of two polarized  $p$ -Barsotti–Tate groups  $H_1 \times H_2$  each of height two and dimension one. The formal deformation ring  $R$  splits as the completed tensor product  $R \cong R_1 \widehat{\otimes}_{\mathbb{F}_p^{\text{alg}}} R_2$  of the formal deformation rings of  $H_1$  and  $H_2$ , and the determinant of the Hasse–Witt matrix of the universal deformation is a pure tensor  $r_1 \otimes r_2$ . If  $H_i \cong (\mathbb{Q}_p/\mathbb{Z}_p) \times \mu_{p^\infty}$  is ordinary then  $R_i \cong \mathbb{F}_p^{\text{alg}}[[x_i]]$  by the theory of Serre–Tate coordinates, and  $r_i \in R_i^\times$ . If  $H_i$  is the unique connected  $p$ -Barsotti–Tate group of dimension one and height two over  $\mathbb{F}_p^{\text{alg}}$  then again Zink’s theory of displays shows that  $R_i \cong \mathbb{F}_p^{\text{alg}}[[x_i]]$  and  $r_i$  is a uniformizing parameter of  $R_i$ . From this we deduce that the quotient of  $R$  by the determinant of the Hasse–Witt matrix has dimension one.  $\square$

Suppose  $\mathcal{F}$  is a coherent  $\mathcal{O}_Y$ -module. We denote by  $\text{cl}(\mathcal{F}) \in \text{CH}_Y^2(M)$  the class determined by  $R\phi_*[\mathcal{F}] \in F^2\mathbf{K}_0^Y(M)$  under the Gillet–Soulé isomorphism (9). Note that the finiteness of  $\phi : Y \rightarrow M$  implies  $R^k\phi_*\mathcal{F} = 0$  for all  $k > 0$ , and so we have simply  $R\phi_*[\mathcal{F}] = [\phi_*\mathcal{F}]$ . As  $\text{cl}(\mathcal{O}_Y)$  is  $H$ -invariant, Lemma 2.1.2 shows that  $\text{cl}(\mathcal{O}_Y)$  arises as the flat pullback of a class

$$\mathcal{C}_p \in \text{CH}_{Y/\mathbb{Z}_p}^2(\mathcal{M}/\mathbb{Z}_p). \tag{28}$$

LEMMA 3.3.2. *The homomorphism (12) takes  $\mathcal{C}_p$  to the cycle  $\mathcal{C}_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{Q}_p$  of (22).*

*Proof.* The construction  $\mathcal{F} \mapsto \text{cl}(\mathcal{F})$  from coherent sheaves to cycle classes is compatible with flat base change, so we may compute the image of  $\mathcal{C}_p$  in

$$\text{CH}_{Y/\mathbb{Q}_p}^2(M/\mathbb{Q}_p) \cong Z_{Y/\mathbb{Q}_p}^2(M/\mathbb{Q}_p)$$

by repeating the construction of  $\mathcal{C}_p$  with  $\mathbb{Z}_p$  replaced by  $\mathbb{Q}_p$ . As  $Y/\mathbb{Q}_p$  is zero-dimensional and reduced, it is a disjoint union of field spectra. One can easily check that the composition

$$\mathbf{K}_0(Y/\mathbb{Q}_p) \xrightarrow{R\phi_*} F^2\mathbf{K}_0^{Y/\mathbb{Q}_p}(M/\mathbb{Q}_p) \rightarrow \text{CH}_{Y/\mathbb{Q}_p}^2(M/\mathbb{Q}_p) \cong Z_{Y/\mathbb{Q}_p}^2(M/\mathbb{Q}_p)$$

takes

$$[\mathcal{F}] \mapsto \sum_{y \in Y/\mathbb{Q}_p} \text{length}_{k(y)}(\mathcal{F}_y) \cdot \phi(y)$$

and the claim follows.  $\square$

DEFINITION 3.3.3. A coherent  $\mathcal{O}_Y$ -module  $\mathcal{F}$  is *skyscraper-free* at a closed point  $y \in Y$  if

$$\text{Hom}_{\mathcal{O}_{Y,y}}(k(y), \mathcal{F}_y) = 0.$$

That is to say,  $y$  is neither an embedded point of  $\mathcal{F}$  nor an irreducible component of  $\text{spt}(\mathcal{F})$ .

LEMMA 3.3.4. *A coherent  $\mathcal{O}_Y$ -module  $\mathcal{F}$  is skyscraper-free at  $y$  if and only if the stalk  $\mathcal{F}_y$  is either trivial or a one-dimensional Cohen–Macaulay  $\mathcal{O}_{Y,y}$ -module.*

*Proof.* Fix a closed point  $y \in Y$  in the support of  $\mathcal{F}$ . By Proposition 3.3.1 the  $\mathcal{O}_{Y,y}$ -module  $\mathcal{F}_y$  has dimension at most one, and hence is Cohen–Macaulay of dimension one if and only if  $\text{depth}(\mathcal{F}_y) \geq 1$ . The condition  $\text{depth}(\mathcal{F}_y) \geq 1$  is equivalent to  $\text{Hom}_{\mathcal{O}_{Y,y}}(k(y), \mathcal{F}_y) = 0$ .  $\square$

By mild abuse of notation we denote again by  $\text{deg}_p$  the composition

$$\text{CH}_{Y_0}^2(M_0) \rightarrow \text{CH}_{\text{ver}}^2(M_0) \xrightarrow{\text{deg}_p} \mathbb{Q}. \tag{29}$$

LEMMA 3.3.5. *Suppose that  $\mathcal{F}$  is a coherent  $\mathcal{O}_Y$ -module which is skyscraper-free at each closed point  $y \in Y_0$ , and assume that the support of the coherent  $\mathcal{O}_{Y_0}$ -module  $j^*\mathcal{F}$  has dimension zero. Then, recalling the pullback  $i^*$  of (10),*

$$\deg_p(i^* \text{cl}(\mathcal{F})) = \sum_{y \in Y_0(\mathbb{F}_p^{\text{alg}})} \text{length}_{\mathcal{O}_{Y_0,y}}(j^*\mathcal{F}_y).$$

*Proof.* Suppose we are given coherent  $\mathcal{O}_M$ -modules  $\mathcal{G}_1, \mathcal{G}_2$  such that  $\mathcal{G}_1 \otimes_{\mathcal{O}_M} \mathcal{G}_2$  is supported in dimension zero. Define the *Serre intersection multiplicity* [Ser00, ch. V] at a closed point  $x \in M$  by

$$I_{\mathcal{O}_{M,x}}^{\text{Serre}}(\mathcal{G}_1, \mathcal{G}_2) = \sum_{\ell \geq 0} (-1)^\ell \text{length}_{\mathcal{O}_{M,x}} \text{Tor}_\ell^{\mathcal{O}_{M,x}}(\mathcal{G}_{1,x}, \mathcal{G}_{2,x}).$$

Lemma 3.3.4 implies that  $\mathcal{F}_y$  is trivial or a one-dimensional Cohen–Macaulay  $\mathcal{O}_{Y,y}$ -module for every closed point  $y \in Y_0$ , and it follows that at every closed point  $x \in M_0$  the stalk  $(\phi_*\mathcal{F})_x$  is either trivial or a one-dimensional Cohen–Macaulay  $\mathcal{O}_{M,x}$ -module. For a closed point  $x \in M_0$  the local ring  $\mathcal{O}_{M_0,x}$  is regular, hence Cohen–Macaulay, and so (abbreviating  $\mathcal{O}_{M_0} = i_*\mathcal{O}_{M_0}$ )  $\mathcal{O}_{M_0,x}$  is a two-dimensional Cohen–Macaulay  $\mathcal{O}_{M,x}$ -module. Applying the corollary of [Ser00, p. 111] we find that

$$\text{Tor}_\ell^{\mathcal{O}_{M,x}}((\phi_*\mathcal{F})_x, \mathcal{O}_{M_0,x}) = 0$$

for all  $\ell > 0$ , and so

$$I_{\mathcal{O}_{M,x}}^{\text{Serre}}(\phi_*\mathcal{F}, \mathcal{O}_{M_0}) = \text{length}_{\mathcal{O}_{M,x}}((\phi_*\mathcal{F})_x \otimes_{\mathcal{O}_{M,x}} \mathcal{O}_{M_0,x}).$$

By examination of the construction of (9), the composition

$$F^2\mathbf{K}_0^{Y_0}(M_0) \rightarrow \text{CH}_{Y_0}^2(M_0) \xrightarrow{\deg_p} \mathbb{Q}$$

is given by

$$[\mathcal{G}] \mapsto \sum_{x \in M_0(\mathbb{F}_p^{\text{alg}})} \text{length}_{\mathcal{O}_{M_0,x}}(\mathcal{G}_x).$$

Putting this all together gives

$$\begin{aligned} \deg_p(i^* \text{cl}(\mathcal{F})) &= \sum_{x \in M_0(\mathbb{F}_p^{\text{alg}})} I_{\mathcal{O}_{M,x}}^{\text{Serre}}(\phi_*\mathcal{F}, \mathcal{O}_{M_0}) \\ &= \sum_{y \in Y_0(\mathbb{F}_p^{\text{alg}})} \text{length}_{\mathcal{O}_{Y_0,y}}(\mathcal{F}_y \otimes_{\mathcal{O}_{M,\phi(y)}} \mathcal{O}_{M_0,\phi(y)}) \end{aligned}$$

and the claim follows. □

We now examine the local rings of  $Y$  at closed points of  $Y_0$ . Let  $W = W(\mathbb{F}_p^{\text{alg}})$  denote the ring of Witt vectors of  $\mathbb{F}_p^{\text{alg}}$ , and denote by **Art** the category of local Artinian  $W$ -algebras with residue field  $\mathbb{F}_p^{\text{alg}}$ . Let  $\mathfrak{G}_0^*$  be a Barsotti–Tate group of dimension one and height two over  $\mathbb{F}_p^{\text{alg}}$ . Thus  $\mathfrak{G}_0^*$  is isomorphic to the  $p$ -divisible group of an elliptic curve. Abbreviate  $\mathcal{O} = \mathcal{O}_F \otimes_{\mathbb{Z}} \mathbb{Z}_p$ . The  $\mathbb{Z}_p$ -basis  $\{\varpi_1, \varpi_2\}$  of  $\mathcal{O}$  determines an injection  $\mathcal{O} \rightarrow M_2(\mathbb{Z}_p)$ , and we define

$$\mathfrak{G}^* = \mathfrak{G}_0^* \times \mathfrak{G}_0^*$$

with the action of  $\mathcal{O}$  induced by the map  $\mathcal{O} \rightarrow M_2(\mathbb{Z}_p)$ . Thus for any  $\mathbb{F}_p^{\text{alg}}$ -algebra  $R$  we have  $\mathfrak{G}^*(R) \cong \mathfrak{G}_0^*(R) \otimes_{\mathbb{Z}_p} \mathcal{O}$ . Let  $\mathbf{D}$  be the functor on **Art** which classifies deformations of  $\mathfrak{G}^*$ . More

precisely, for an object  $R$  of  $\mathbf{Art}$  let  $\mathbf{D}(R)$  be the set of isomorphism classes of pairs  $(\mathfrak{G}, \rho)$  in which  $\mathfrak{G}$  is a Barsotti–Tate group over  $R$  with an action of  $\mathcal{O}$  and

$$\rho : \mathfrak{G}^* \rightarrow \mathfrak{G} /_{\mathbb{F}_p^{\text{alg}}}$$

is an isomorphism of Barsotti–Tate groups over  $\mathbb{F}_p^{\text{alg}}$  respecting  $\mathcal{O}$ -actions. Let  $\tau \in \text{End}(\mathfrak{G}^*)$  be an endomorphism of  $\mathfrak{G}^*$  which commutes with the action of  $\mathcal{O}$ . For each object  $R$  of  $\mathbf{Art}$  let  $\mathbf{D}_\tau(R) \subset \mathbf{D}(R)$  to be the subset consisting of those deformations for which the endomorphism  $\tau$  lifts: that is to say, those deformations  $(\mathfrak{G}, \rho) \in \mathbf{D}(R)$  for which there exists a (necessarily unique)  $\tilde{\tau} \in \text{End}(\mathfrak{G})$  which commutes with the action of  $\mathcal{O}$  and satisfies

$$\tilde{\tau} \circ \rho = \rho \circ \tau$$

as elements of  $\text{Hom}(\mathfrak{G}^*, \mathfrak{G} /_{\mathbb{F}_p^{\text{alg}}})$ .

**PROPOSITION 3.3.6.** *The functor  $\mathbf{D}$  is pro-represented by a  $W$ -algebra  $R^{\text{univ}}$  isomorphic to a power series ring in two variables over  $W$ , and  $\mathbf{D}_\tau$  is pro-represented by  $R^{\text{univ}}/J_\tau$  for some ideal  $J_\tau \subset R^{\text{univ}}$ .*

*Proof.* The first claim follows from work of Rapoport [Rap78] and Deligne and Pappas [DP94], who have determined the non-smooth locus of Hilbert–Blumenthal schemes. Briefly, by the Serre–Tate theorem the functor  $\mathbf{D}$  is represented by the completion of the local ring of a Hilbert–Blumenthal scheme of relative dimension two over  $W$ . The smoothness of the local ring at this point is a consequence of the fact that

$$\text{Lie}(\mathfrak{G}^*) \cong \text{Lie}(\mathfrak{G}_0^*) \otimes_{\mathbb{Z}_p} \mathcal{O}$$

is free of rank one over  $\mathbb{F}_p^{\text{alg}} \otimes_{\mathbb{Z}_p} \mathcal{O}$ , and so satisfies the *Rapoport condition* of [Vol05]. See especially [Vol05, Remark 3.8]. The representability of  $\mathbf{D}_\tau$  follows from [RZ96, Proposition 2.9].  $\square$

**PROPOSITION 3.3.7.** *Suppose that  $\mathfrak{G}_0^*$  is isomorphic to the Barsotti–Tate group of an ordinary elliptic curve. Then the ideal  $J_\tau \subset R^{\text{univ}}$  can be generated by two elements.*

*Proof.* This follows from the theory of Serre–Tate coordinates as in [Gor02, Theorem 4.2] or [Hid04, Theorem 8.9]. If we define rank one  $\mathcal{O}$ -modules

$$P = \varprojlim \mathfrak{G}^*(\mathbb{F}_p^{\text{alg}})[p^k], \quad Q = \varprojlim \text{Hom}(\mathfrak{G}^*, \mu_{p^k})$$

then there is a canonical isomorphism of functors on  $\mathbf{Art}$ ,

$$\mathbf{D} \cong \text{Hom}_{\mathbb{Z}_p}(P \otimes_{\mathcal{O}} Q, \mu_{p^\infty}).$$

The endomorphism  $\tau$  of  $\mathfrak{G}$  induces an  $\mathcal{O}$ -linear endomorphism  $x \mapsto t_P(x)$  of  $P$  and an  $\mathcal{O}$ -linear endomorphism  $x \mapsto t_Q(x)$  of  $Q$ , and there is an isomorphism

$$\mathbf{D}_\tau \cong \text{Hom}_{\mathbb{Z}_p}((P \otimes_{\mathcal{O}} Q)/I, \mu_{p^\infty})$$

where  $I$  is the  $\mathbb{Z}_p$ -submodule of  $P \otimes_{\mathcal{O}} Q$ ,

$$I = \{t_P(x) \otimes y - x \otimes t_Q(y) \mid x \in P, y \in Q\}.$$

After choosing an appropriate  $\mathbb{Z}_p$ -basis of  $P \otimes_{\mathcal{O}} Q$  we find that  $\mathbf{D}_\tau \cong \mu_{p^s} \times \mu_{p^t}$  for some  $0 \leq s, t \leq \infty$ . The claim follows.  $\square$

**PROPOSITION 3.3.8.** *Suppose that  $\mathfrak{G}_0^*$  is isomorphic to the Barsotti–Tate group of a supersingular elliptic curve. Then the ideal  $J_\tau \subset R^{\text{univ}}$  can be generated by two elements.*

*Proof.* We modify the argument of [Wew07, Proposition 5.1], and use Zink’s theory of displays [Zin02] in order to avoid the messy language of formal group laws (and formal group cohomology) in dimension two. Denote by  $\mathfrak{G}^{\text{univ}}$  the universal deformation of  $\mathfrak{G}^*$  over  $R^{\text{univ}}$ , and let  $\mathfrak{m}$  denote the maximal ideal of  $R^{\text{univ}}$ . Thus the endomorphism  $\tau$  of  $\mathfrak{G}^*$  lifts to an  $\mathcal{O}$ -linear endomorphism of  $\mathfrak{G}^{\text{univ}}$  over  $R^{\text{univ}}/J_\tau$  (which we again denote by  $\tau$ ), but not over

$$R = R^{\text{univ}}/\mathfrak{m}J_\tau.$$

Let  $\mathfrak{a} = J_\tau/\mathfrak{m}J_\tau$  be the kernel of the surjection

$$R \rightarrow R^{\text{univ}}/J_\tau$$

so that  $\mathfrak{m} \cdot \mathfrak{a} = 0$ . To the Barsotti–Tate group  $\mathfrak{G}^{\text{univ}}$  over  $R^{\text{univ}}$  Zink’s theory attaches a universal display  $(P, Q, F, V^{-1})^{\text{univ}}$  over  $R^{\text{univ}}$ , and we denote by  $(P, Q, F, V^{-1})$  the reduction of the universal display to  $R$ . For our purposes, we need only know that  $P$  is a free module of rank four over the Witt vectors  $W(R)$  with an action of  $\mathcal{O}$ , that  $Q \subset P$  is an  $\mathcal{O} \otimes_{\mathbb{Z}_p} W(R)$ -submodule with the property that  $P/Q$  is annihilated by the kernel  $I_R$  of  $W(R) \rightarrow R$ , and that

$$Q/I_R P \quad \text{and} \quad P/Q$$

are each free of rank two over  $R$ . We claim that, in addition,  $Q/I_R P$  is generated as an  $\mathcal{O} \otimes_{\mathbb{Z}_p} R$ -module by a single element. Indeed, let  $(P_0^*, Q_0^*, F_0^*, V_0^{-1*})$  and  $(P^*, Q^*, F^*, V^{-1*})$  be the displays of  $\mathfrak{G}_0^*$  and  $\mathfrak{G}^*$ , respectively. Then, letting  $I_{\mathbb{F}_p^{\text{alg}}}$  denote the kernel of  $W \rightarrow \mathbb{F}_p^{\text{alg}}$ , we have

$$(Q/I_R P) \otimes_R \mathbb{F}_p^{\text{alg}} \cong Q^*/I_{\mathbb{F}_p^{\text{alg}}} P^* \cong (Q_0^*/I_{\mathbb{F}_p^{\text{alg}}} P_0^*) \otimes_{\mathbb{Z}_p} \mathcal{O}.$$

As  $Q_0^*/I_{\mathbb{F}_p^{\text{alg}}} P_0^*$  is free of rank one over  $\mathbb{F}_p^{\text{alg}}$ , we see that  $(Q/I_R P) \otimes_R \mathbb{F}_p^{\text{alg}}$  is free of rank one over  $\mathcal{O} \otimes_{\mathbb{Z}_p} \mathbb{F}_p^{\text{alg}}$ . By Nakayama’s lemma it follows that  $Q/I_R P$  is generated by a single element over  $\mathcal{O} \otimes_{\mathbb{Z}_p} R$ .

By Zink’s theory [Zin02, Definition 72] the obstruction to lifting the endomorphism  $\tau$  from  $R^{\text{univ}}/J_\tau$  to  $R$  is given by a nontrivial homomorphism of  $\mathcal{O} \otimes_{\mathbb{Z}_p} R$ -modules,

$$\text{Obst} : Q/I_R P \rightarrow \mathfrak{a} \otimes_R P/Q.$$

Let  $\gamma$  generate  $Q/I_R P$  as an  $\mathcal{O} \otimes_{\mathbb{Z}_p} R$ -module. If we pick an  $R$ -module basis  $\{e_1, e_2\} \subset P/Q$  then

$$\text{Obst}(\gamma) = a_1 \otimes e_1 + a_2 \otimes e_2$$

for some  $a_1, a_2 \in \mathfrak{a}$ , and the composition

$$Q/I_R P \xrightarrow{\text{Obst}} \mathfrak{a} \otimes_R P/Q \rightarrow (\mathfrak{a}/(a_1, a_2)) \otimes_R P/Q$$

is the trivial map. This implies [Zin02, (119)] that  $\tau$  can be lifted from an endomorphism over  $R^{\text{univ}}/J_\tau$  to an endomorphism over  $R/(a_1, a_2)$ , and therefore  $(a_1, a_2) = \mathfrak{a}$ . Thus  $J_r/\mathfrak{m}J_r$  is generated as an  $R^{\text{univ}}$ -module by two elements, and so Nakayama’s lemma implies that the ideal  $J_r$  is also generated by two elements. □

**COROLLARY 3.3.9.** *Let  $y$  be a closed point of  $Y_0$ . The structure sheaf  $\mathcal{O}_Y$  is skyscraper-free at  $y$ .*

*Proof.* As above let  $W$  denote the ring of Witt vectors of  $\mathbb{F}_p^{\text{alg}}$  and choose a point  $z \in Y/W$  lying above  $y$ . By Lemma 3.3.4 and standard results in commutative algebra (e.g. [Mat89, Theorem 17.5 and the corollary to Theorem 23.3]) it suffices to prove that the completed local ring  $\mathcal{O}_{Y/W, z}^\wedge$  is Cohen–Macaulay of dimension one. The geometric point  $\text{Spec}(k(z)) \rightarrow Y/W$  corresponds to a

triple  $(\mathbf{A}, \lambda, t_\alpha)$  with  $\mathbf{A} = \mathbf{A}_0 \otimes \mathcal{O}_F$  for some QM abelian surface  $\mathbf{A}_0$  over  $k(z) = \mathbb{F}_p^{\text{alg}}$ . Using the splitting  $\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p \cong M_2(\mathbb{Z}_p)$  the Barsotti–Tate group of  $\mathbf{A}_0$  splits as  $\mathfrak{G}_0^* \times \mathfrak{G}_0^*$ , and the Barsotti–Tate group of  $\mathbf{A}$  splits as  $\mathfrak{G}^* \times \mathfrak{G}^*$ . As the Barsotti–Tate group of any deformation of  $\mathbf{A}$  admits a similar splitting into two isomorphic deformations of  $\mathfrak{G}^*$ , it follows from the Serre–Tate theorem that the formal deformation space of the pair  $(\mathbf{A}, \lambda)$  is isomorphic to  $\text{Spf}(R^{\text{univ}})$ , while the formal deformation space of the triple  $(\mathbf{A}, \lambda, t_\alpha)$  is isomorphic to  $\text{Spf}(R_\tau^{\text{univ}})$  for a suitable  $\tau \in \text{End}_{\mathcal{O}}(\mathfrak{G}^*)$  (the polarization  $\lambda$  lifts uniquely to any deformation of  $\mathbf{A}$  by the Corollary to [Vol05, Theorem 3] or by [BBGGLMS79, p. 51]). In other words  $\mathcal{O}_{Y/W,z}^\wedge \cong R_\tau^{\text{univ}}$ . Thus Propositions 3.3.6, 3.3.7, and 3.3.8 imply that

$$\mathcal{O}_{Y/W,z}^\wedge \cong W[[s_1, s_2]]/(f_1, f_2) \tag{30}$$

for some power series  $f_1, f_2 \in W[[s_1, s_2]]$ . We know from Proposition 3.3.1 that the ring on the left-hand side of (30) has dimension at most one, and it follows from (30) that  $\mathcal{O}_{Y/W,z}^\wedge$  is Cohen–Macaulay of dimension exactly one (by, for example, [Liu02, Exercise 6.3.4 and Corollary 8.2.18]).  $\square$

PROPOSITION 3.3.10. *If  $Y_0$  is zero-dimensional then the class  $C_p$  satisfies*

$$\deg_p(i^* C_p) = \sum_{y \in Y_0(\mathbb{F}_p^{\text{alg}})} \text{length}_{\mathcal{O}_{Y_0,y}}(\mathcal{O}_{Y_0,y}).$$

*Proof.* Set  $\mathcal{F} = \mathcal{O}_Y$  in Lemma 3.3.5 and combine with Corollary 3.3.9.  $\square$

#### 4. Construction of cycles at primes of bad reduction

Throughout §4 we fix a prime  $p$  at which  $B_0$  is ramified, and a totally positive  $\alpha \in \mathcal{O}_F$ . By Hypothesis A the prime  $p$  splits in  $\mathcal{O}_F$ , say as  $p\mathcal{O}_F = \mathfrak{p}_1\mathfrak{p}_2$ . Our goal is to construct and examine a cycle class

$$C_p \in \text{CH}_{\mathcal{Y}(\alpha)/\mathbb{Z}_p}^2(\mathcal{M}/\mathbb{Z}_p).$$

The definition used in (28) breaks down for  $p \mid \text{disc}(B_0)$  as the algebraic stack  $\mathcal{Y}(\alpha)_{/\mathbb{Z}_p}$  may have vertical components of dimension two. Our modified construction will make use of calculations of Kudla and Rapoport [KR00] (and of Kudla, Rapoport and Yang [KRY04, KRY06] for the case  $p = 2$ ).

The following notation will be used throughout §4. Abbreviate  $\mathcal{Y} = \mathcal{Y}(\alpha)$  and  $\mathcal{Y}_0 = \mathcal{Y}_0(\alpha)$ . Let  $W = W(\mathbb{F}_p^{\text{alg}})$  be the ring of Witt vectors of  $\mathbb{F}_p^{\text{alg}}$ . Denote by  $\mathbf{Nilp}_{\mathbb{Z}_p}$  (respectively  $\mathbf{Nilp}_W$ ) the category of  $\mathbb{Z}_p$ -schemes (respectively  $W$ -schemes) on which  $p$  is locally nilpotent. Let  $\mathbb{A}_f^p$  denote the prime-to- $p$  finite adèles of  $\mathbb{Q}$ . Let  $U \subset U^{\text{max}}$  be a normal compact open subgroup which factors as  $U = U_p U^p$  with  $U_p = U_p^{\text{max}}$  and  $U^p \subset G(\mathbb{A}_f^p)$ . Let  $M$  denote the algebraic stack over  $\mathbb{Z}_p$  whose objects are triples  $(\mathbf{A}, \lambda, \nu)$  in which  $(\mathbf{A}, \lambda)$  is a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold over a  $\mathbb{Z}_p$ -scheme  $S$  and  $\nu$  is a level  $U$  structure on  $(\mathbf{A}, \lambda)$ . Define  $Y, M_0,$  and  $Y_0,$  by (14) and assume that  $U^p$  is chosen small enough that the algebraic stack  $M$  (and hence also  $Y, M_0,$  and  $Y_0$ ) is a scheme. Set  $H = U^{\text{max}}/U$  so that  $\mathcal{M}/\mathbb{Z}_p \cong [H \backslash M]$ . Let  $\widehat{M}$  and  $\widehat{Y}$  denote the formal completions

of  $M$  and  $Y$  along their special fibers. As in § 3.1 we have the cartesian diagram

$$\begin{array}{ccc} \mathcal{Y}_0 & \xrightarrow{\phi_0} & \mathcal{M}_0 \\ j \downarrow & & \downarrow i \\ \mathcal{Y} & \xrightarrow{\phi} & \mathcal{M} \end{array}$$

and similarly with  $\mathcal{M}, \mathcal{M}_0, \mathcal{Y}$ , and  $\mathcal{Y}_0$  replaced by  $M, M_0, Y$ , and  $Y_0$ .

### 4.1 The Cerednik–Drinfeld uniformization

We first describe some simple generalizations of the Cerednik–Drinfeld uniformization of  $M_0$ . As in § 3.1 let

$$\widehat{\Lambda}_0^p = \prod_{\ell \neq p} (\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_\ell)$$

be the restricted topological product, and define  $\widehat{\Lambda}^p$  in the same way, replacing  $B_0$  by  $B$ . Fix a principally polarized QM abelian surface  $(\mathbf{A}_0^*, \lambda_0^*)$  over  $\mathbb{F}_p^{\text{alg}}$ , let  $A_0^*$  be the abelian surface underlying  $\mathbf{A}_0^*$ , and choose an isomorphism

$$\nu_0^* : \widehat{\Lambda}_0^p \rightarrow \text{Ta}^p(A_0^*)$$

of left  $\widehat{\mathcal{O}}_{B_0}^p$ -modules, where  $\text{Ta}^p(A_0^*)$  is the prime-to- $p$  adelic Tate module of  $A_0^*$ , in such a way that the Weil pairing

$$\text{Ta}^p(A_0) \times \text{Ta}^p(A_0) \rightarrow \widehat{\mathbb{Z}}^p(1)$$

induced by  $\lambda_0^*$  agrees with the pairing (19) up to a  $(\widehat{\mathbb{Z}}^p)^\times$ -multiple. Define a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold

$$(\mathbf{A}^*, \lambda^*) = (\mathbf{A}_0^* \otimes \mathcal{O}_F, \lambda_0^* \otimes \mathcal{O}_F).$$

Letting  $A^*$  denote the abelian fourfold underlying  $\mathbf{A}^*$ ,  $\nu_0^*$  induces an isomorphism

$$\nu^* : \widehat{\Lambda}^p \rightarrow \text{Ta}^p(A^*)$$

of left  $\widehat{\mathcal{O}}_B^p$ -modules which is a  $U$ -level structure on  $(\mathbf{A}^*, \lambda^*)$ . Define totally definite quaternion algebras over  $\mathbb{Q}$  and  $F$ , respectively,

$$\overline{B}_0 = \text{End}^0(\mathbf{A}_0^*), \quad \overline{B} = \text{End}^0(\mathbf{A}^*).$$

Let  $\overline{G}_0 \subset \overline{G}$  be the algebraic groups over  $\mathbb{Q}$  defined in the same way as  $G_0 \subset G$ , but with  $B_0$  and  $B$  replaced by  $\overline{B}_0$  and  $\overline{B}$ .

Let  $\mathfrak{G}_0^*$  denote the  $p$ -divisible group of  $A_0^*$  equipped with its action of  $\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ . For any object  $S$  of  $\mathbf{Nilp}_W$  with special fiber  $S_{/\mathbb{F}_p^{\text{alg}}}$ , denote by  $\mathfrak{h}_m(S)$  the set of isomorphism classes of pairs  $(\mathfrak{G}_0, \rho_0)$  in which  $\mathfrak{G}_0$  is a special formal  $\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ -module of dimension two and height four over  $S$  in the sense of [BC92, § II.2], and

$$\rho_0 \in \text{Hom}(\mathfrak{G}_0^* \times_{\mathbb{F}_p^{\text{alg}}} S_{/\mathbb{F}_p^{\text{alg}}}, \mathfrak{G}_0 \times_S S_{/\mathbb{F}_p^{\text{alg}}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$$

is a height  $2m$  quasi-isogeny of  $p$ -divisible groups over  $S_{/\mathbb{F}_p^{\text{alg}}}$  respecting the action of  $\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ . By a theorem of Drinfeld [BC92, Théorème II.8.2] there is a formal  $W$ -scheme  $\mathfrak{h}_m$  whose functor of points is  $S \mapsto \mathfrak{h}_m(S)$ , and we set

$$\mathfrak{X}_0 = \bigsqcup_{m \in \mathbb{Z}} \mathfrak{h}_m.$$



The group  $\overline{G}_0(\mathbb{Q}_p)$  acts on  $\mathfrak{X}_0$  by

$$\gamma \cdot (\mathfrak{G}_0, \rho_0) = (\mathfrak{G}_0, \rho_0 \circ \gamma^{-1}),$$

taking the component  $\mathfrak{h}_m$  to the component  $\mathfrak{h}_{m-\text{ord}_p \text{Nm}(\gamma)}$ . The isomorphism (16) provides an isomorphism  $\overline{B}_0 \otimes_{\mathbb{Q}} F \cong \overline{B}$ , and so, as  $p$  splits in  $F$ , an isomorphism

$$\overline{G}(\mathbb{Q}_p) \cong \{(x, y) \in \overline{G}_0(\mathbb{Q}_p) \times \overline{G}_0(\mathbb{Q}_p) \mid \text{Nm}(x) = \text{Nm}(y)\}.$$

This determines an action of  $\overline{G}(\mathbb{Q}_p)$  on

$$\mathfrak{X} = \bigsqcup_{m \in \mathbb{Z}} (\mathfrak{h}_m \times_W \mathfrak{h}_m).$$

There is a unique isomorphism

$$\iota_0 : \overline{B}_0 \otimes_{\mathbb{Q}} \mathbb{A}_f^p \rightarrow B_0 \otimes_{\mathbb{Q}} \mathbb{A}_f^p$$

of  $\mathbb{A}_f^p$ -modules such that for every prime  $\ell \neq p$  and every  $g \in \overline{B}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$  the diagram

$$\begin{array}{ccc} B_0 \otimes_{\mathbb{Q}} \mathbb{Q}_\ell & \xrightarrow{\nu_0^*} & \text{Ta}_\ell(A_0^*) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell \\ \cdot \iota_0(g) \downarrow & & \downarrow g \\ B_0 \otimes_{\mathbb{Q}} \mathbb{Q}_\ell & \xrightarrow{\nu_0^*} & \text{Ta}_\ell(A_0^*) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell \end{array}$$

commutes, where the vertical arrow on the left is  $x \mapsto x \cdot \iota_0(g)$ . The function  $\iota_0$  satisfies  $\iota_0(gh) = \iota_0(h)\iota_0(g)$ . Let

$$\iota : \overline{B} \otimes_{\mathbb{Q}} \mathbb{A}_f^p \rightarrow B \otimes_{\mathbb{Q}} \mathbb{A}_f^p$$

be the  $F \otimes_{\mathbb{Q}} \mathbb{A}_f^p$ -module map obtained by tensoring  $\iota_0$  with  $F$ . There are induced isomorphisms  $\iota_0 : \overline{G}_0(\mathbb{A}_f^p) \rightarrow G_0^{\text{op}}(\mathbb{A}_f^p)$  and  $\iota : \overline{G}(\mathbb{A}_f^p) \rightarrow G^{\text{op}}(\mathbb{A}_f^p)$  where op denotes the opposite group. For any compact open subgroup  $C \subset G(\mathbb{A}_f^p)$  define

$$\overline{C} = \iota^{-1}(C) \subset \overline{G}(\mathbb{A}_f^p).$$

PROPOSITION 4.1.1. *There is an isomorphism of formal  $W$ -schemes*

$$\widehat{M}/_W \cong \overline{G}(\mathbb{Q}) \backslash \mathfrak{X} \times \overline{G}(\mathbb{A}_f^p) / \overline{U}^p.$$

*Proof.* This follows from the general  $p$ -adic uniformization results of [Bou97, BZ, RZ96], although in this simple case the proof is a straightforward imitation of the proof of the Cerednik–Drinfeld uniformization as in [BC92]. We describe the isomorphism on  $\mathbb{F}_p^{\text{alg}}$ -valued points. First let us show that there is a single  $\mathcal{O}_B$ -linear isogeny class of QM abelian fourfolds over  $\mathbb{F}_p^{\text{alg}}$ . Let  $\mathbf{A} = (A, i)$  be a QM abelian fourfold over  $\mathbb{F}_p^{\text{alg}}$  with  $p$ -divisible group  $\mathfrak{G} = A[p^\infty]$ . As  $p$  splits in  $F$  there is a decomposition  $\mathfrak{G} \cong \mathfrak{G}^1 \oplus \mathfrak{G}^2$  in which  $\mathcal{O}_F$  acts on  $\mathfrak{G}^i$  through the map  $\mathcal{O}_F \rightarrow \mathcal{O}_{F, \mathfrak{p}_i} \cong \mathbb{Z}_p$ , and a corresponding decomposition of the Lie algebra

$$\text{Lie}(A) \cong \text{Lie}(\mathfrak{G}^1) \oplus \text{Lie}(\mathfrak{G}^2).$$

The Kottwitz condition on the action of  $\mathcal{O}_B$  on  $\text{Lie}(A)$  implies that each  $\mathfrak{G}^i$  has dimension two and is a special formal  $\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ -module in the sense of [BC92, § II.2]. By [BC92, Proposition II.5.1] each  $\mathfrak{G}^i$  has height four, and so is isogenous to the square of the  $p$ -divisible group of a supersingular elliptic curve by [BC92, § III.4]. Now let  $\mathbb{F}$  be a finite field of  $q = p^k$  elements chosen large enough that  $A$ , all simple isogeny factors of  $A$ , and all endomorphisms of  $A$  are

defined over  $\mathbb{F}$ . Let  $A'$  be a simple isogeny factor of  $A$  defined over  $\mathbb{F}$  and let  $\pi \in \text{End}(A')$  be the  $q$ -power Frobenius endomorphism. As the slopes of the  $p$ -divisible group of  $A'$  are all  $1/2$ , the argument of [BC92, Proposition III.2] shows that  $\pi^2/q$  is a root of unity in the field  $\mathbb{Q}(\pi)$ , and so by enlarging the field  $\mathbb{F}$  we have  $\pi^2 = q$ . By the Honda–Tate theory  $A'$  is a supersingular elliptic curve. Thus  $A$  is isogenous to the fourth power of a supersingular elliptic curve, and so any two QM abelian fourfolds are isogenous. The argument of [Mil79, Remark 5.3] shows that this isogeny can be chosen to be  $\mathcal{O}_B$ -linear.

Now suppose we are given a  $\mathfrak{D}^{-1}$ -polarized QM abelian fourfold  $(\mathbf{A}, \lambda)$  over  $\mathbb{F}_p^{\text{alg}}$ . The claim is that an  $\mathcal{O}_B$ -linear quasi-isogeny  $\rho \in \text{Hom}(\mathbf{A}^*, \mathbf{A}) \otimes_{\mathcal{O}_F} F$  can be chosen in such a way that  $\rho^\vee \circ \lambda \circ \rho = \lambda^*$ . Arguing as in [BC92, § III.4] or [Vol05, Proposition 1.3], the polarization module  $\mathcal{P}$  of  $\mathbf{A}^*$ , defined as the  $\mathcal{O}_F$ -module of all  $\phi \in \text{Hom}(A^*, A^{*\vee})$  such that

- (a)  $\phi$  is  $\mathcal{O}_B$ -linear,
- (b)  $\phi = \phi^\vee$ ,
- (c)  $\phi \circ i(b^*) = i(b)^\vee \circ \phi$  for all  $b \in \mathcal{O}_B$ ,

is projective of rank one, with the subset of polarizations  $\mathcal{P}^+ \subset \mathcal{P}$  forming a positive cone (this is also stated without proof on [BBGGLMS79, p. 54]). Therefore,  $\lambda^*$  and  $\rho^\vee \circ \lambda \circ \rho$  differ by the action of the group of totally positive elements of  $F^\times$ . Write  $\rho^\vee \circ \lambda \circ \rho = \lambda^* \circ i(\beta)$  with  $\beta \in F^\times$  totally positive. Viewing  $\overline{B}$  as an  $F$ -algebra we may pick  $b \in \overline{B}$  of reduced norm  $\beta$ . As  $\overline{B}$  is the centralizer of  $B$  in  $\text{End}(A^*) \otimes_{\mathbb{Z}} \mathbb{Q}$ , it is stable under the Rosati involution determined by  $\lambda^*$ , and the Rosati involution induces the main involution on  $\overline{B}$  (the only positive involution of a totally definite quaternion algebra). It follows that

$$\rho^\vee \circ \lambda \circ \rho = b^\vee \circ \lambda^* \circ b$$

and so replacing  $\rho$  by  $\rho \circ b^{-1}$  gives  $\rho^\vee \circ \lambda \circ \rho = \lambda^*$ .

We now make the isomorphism of the proposition explicit on geometric points. Fix a triple  $(\mathbf{A}, \lambda, \nu) \in M(\mathbb{F}_p^{\text{alg}})$  and (denoting by  $A$  the abelian scheme underlying  $\mathbf{A}$ ) let  $\rho: A^* \rightarrow A$  be an  $\mathcal{O}_B$ -linear isogeny chosen so that  $\rho^\vee \circ \lambda \circ \rho$  is an integer multiple of  $\lambda^*$ . Attached to  $\rho: A^* \rightarrow A$  is an  $\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p$ -linear isogeny of  $p$ -divisible groups  $\rho: \mathfrak{G}^* \rightarrow \mathfrak{G}$  which, using the splittings

$$\mathfrak{G}^* \cong \mathfrak{G}_0^* \times \mathfrak{G}_0^*, \quad \mathfrak{G} \cong \mathfrak{G}^1 \times \mathfrak{G}^2$$

induced by  $\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p \cong (\mathcal{O}_{B_0} \otimes_{\mathbb{Z}} \mathbb{Z}_p)^2$ , determines an element

$$(\mathfrak{G}, \rho) = ((\mathfrak{G}^1, \rho^1), (\mathfrak{G}^2, \rho^2)) \in \mathfrak{X}_0(\mathbb{F}_p^{\text{alg}}) \times \mathfrak{X}_0(\mathbb{F}_p^{\text{alg}}).$$

As the polarizations  $\lambda$  and  $\lambda^*$  each have degree prime to  $p$ , the equality  $\rho^\vee \circ \lambda \circ \rho = m\lambda^*$  with  $m \in \mathbb{Z}$  implies that  $\deg(\rho^1) = \deg(\rho^2)$ , and so the pair  $(\mathfrak{G}, \rho)$  lies in  $\mathfrak{X}(\mathbb{F}_p^{\text{alg}})$ . There is a unique (up to right multiplication by  $\overline{U}$ )  $g \in \overline{G}(\mathbb{A}_f^p)$  for which the diagram

$$\begin{array}{ccc} \widehat{\Lambda}^p & \xrightarrow{\nu^*} & \text{Ta}^p(A^*) \\ \downarrow \cdot \iota(g^{-1}) & & \downarrow \rho \\ \widehat{\Lambda}^p & \xrightarrow{\nu} & \text{Ta}^p(A) \end{array}$$

commutes. The isomorphism of the proposition then takes the triple  $(\mathbf{A}, \lambda, \nu)$  to the double coset of  $((\mathfrak{G}, \rho), g)$ .

The inverse function is constructed as follows. Starting from  $((\mathfrak{G}, \rho), g)$  we may act on the left by an element of  $\overline{G}(\mathbb{Q})$  to assume that  $\rho: \mathfrak{G}^* \rightarrow \mathfrak{G}$  is an isogeny (as opposed to merely a quasi-

isogeny) and that left multiplication by  $\iota(g^{-1})$  stabilizes  $\widehat{\Lambda}^p$ . There is then a unique choice of QM abelian fourfold  $\mathbf{A} = (A, i)$  over  $\mathbb{F}_p^{\text{alg}}$ ,  $\widehat{\mathcal{O}}_B^p$ -linear isomorphism  $\nu : \widehat{\Lambda}^p \rightarrow \text{Ta}^p(A)$ , isomorphism  $A[p^\infty] \cong \mathfrak{G}$ , and extension of  $\rho$  to an isogeny  $\rho : \mathbf{A}^* \rightarrow \mathbf{A}$  for which the above diagram commutes. Define a quasi-polarization  $\lambda \in \text{Hom}(A, A^\vee) \otimes_{\mathbb{Z}} \mathbb{Q}$  by  $\rho^\vee \circ \lambda \circ \rho = \lambda^*$ . The claim is that there is a positive integer multiple of  $\lambda$  which is a  $\mathfrak{D}^{-1}$ -polarization of  $\mathbf{A}$ . If such a multiple exists, it is clearly unique. First pick any positive integer  $m$  large enough that  $m\lambda$  is a polarization, and consider the induced polarization  $m\lambda : \mathfrak{G} \rightarrow \mathfrak{G}^*$ . Using the splitting  $\mathfrak{G} \cong \mathfrak{G}^1 \times \mathfrak{G}^2$  and [BC92, Lemme III.4.2] we see that the kernel of  $m\lambda$  has the form  $\mathfrak{G}^1[p^{a_1}] \times \mathfrak{G}^2[p^{a_2}]$  for some integers  $a_1$  and  $a_2$ , but from the fact that  $\rho^1$  and  $\rho^2$  have the same degree one deduces that  $a_1 = a_2$ . Thus after dividing  $m$  by a power of  $p$  we may assume that  $m\lambda$  is a polarization of degree prime to  $p$ . Recalling that the pairing

$$\psi : \widehat{\Lambda}^p \times \widehat{\Lambda}^p \rightarrow \widehat{\mathbb{Z}}^p$$

of § 3.1 agrees with the pairing induced by  $\nu^*$ , the isogeny  $\lambda^*$ , and the Weil pairing on  $\text{Ta}^p(A^*)$ , one can show that the pairing on  $\widehat{\Lambda}^p$  induced by  $\nu$ , the polarization  $m\lambda$ , and the Weil pairing on  $\text{Ta}^p(A)$  is exactly  $m \cdot \text{Nm}(g) \cdot \psi$ . Choosing a positive integer  $k$  prime to  $p$  and satisfying  $\text{ord}_\ell(k) = \text{ord}_\ell(\text{Nm}(g^{-1}))$  for every prime  $\ell \neq p$ , we deduce that  $(m/k)\mathfrak{D} \subset \mathcal{O}_F$  and that  $m\lambda$  has kernel  $A[(m/k)\mathfrak{D}]$ . It follows that  $k\lambda$  is a  $\mathfrak{D}^{-1}$ -polarization of  $\mathbf{A}$ .  $\square$

DEFINITION 4.1.2. Suppose  $\Sigma$  is a finite set of prime ideals of  $\mathcal{O}_F$  and let

$$\mathcal{O}_{F,\Sigma} = \{x \in F \mid \text{ord}_{\mathfrak{q}}(x) \geq 0 \forall \mathfrak{q} \notin \Sigma\}$$

be the ring of  $\Sigma$ -integers in  $\mathcal{O}_F$ . If  $\mathbf{A}$  is a QM abelian fourfold over a  $\mathbb{Z}_p$ -scheme  $S$  and  $\tau \in \text{End}^0(\mathbf{A})$ , we say that  $\tau$  is *integral away from  $\Sigma$*  if  $\tau$  lies in the image of the inclusion

$$\text{End}(\mathbf{A}) \otimes_{\mathcal{O}_F} \mathcal{O}_{F,\Sigma} \hookrightarrow \text{End}^0(\mathbf{A}).$$

Define set-valued functors on  $\mathbf{Nilp}_W$  as follows. For  $S$  an object of  $\mathbf{Nilp}_W$  let  $\mathfrak{M}(S)$  be the set of isomorphism classes of quadruples  $(\mathbf{A}, \lambda, \nu, t_\alpha)$  in which  $(\mathbf{A}, \lambda, \nu) \in M(S)$  and  $t_\alpha \in \text{End}^0(\mathbf{A})$  is trace zero, integral away from  $\{\mathfrak{p}_1, \mathfrak{p}_2\}$ , and satisfies  $Q(t_\alpha) = \alpha$ . For  $k \in \{1, 2\}$  we let  $\mathfrak{M}^k(S) \subset \mathfrak{M}(S)$  be the subset of quadruples  $(\mathbf{A}, \lambda, \nu, t_\alpha)$  for which  $t_\alpha$  is integral away from  $\mathfrak{p}_k$ . There are evident morphisms

$$\begin{array}{ccccc}
 & & \mathfrak{M}^1 & & \\
 & \nearrow & & \searrow & \\
 \widehat{Y}_{/W} & & & & \mathfrak{M} \longrightarrow \widehat{M}_{/W} \\
 & \searrow & & \nearrow & \\
 & & \mathfrak{M}^2 & & 
 \end{array} \tag{31}$$

and the square is cartesian. Fix a connected object  $S$  of  $\mathbf{Nilp}_W$  and let  $\overline{V}_0$  and  $\overline{V}$  be trace zero elements of  $\overline{B}_0$  and  $\overline{B}$ , respectively. If for each  $\tau \in \overline{V}$  we define

$$\Omega(\tau) = \{g \in \overline{G}(\mathbb{A}_f^p) \mid \widehat{\Lambda}^p \cdot \iota(g^{-1}\tau g) \subset \widehat{\Lambda}^p\}$$

then Proposition 4.1.1 generalizes to a uniformization

$$\mathfrak{M} \cong \overline{G}(\mathbb{Q}) \backslash \bigsqcup_{\substack{\tau \in \overline{V} \\ \overline{Q}(\tau) = \alpha}} (\mathfrak{X} \times \Omega(\tau) / \overline{U}^p) \tag{32}$$

where  $\overline{Q}$  is the quadratic form  $\overline{Q}(\tau) = -\tau^2$ . On geometric points the isomorphism is defined as follows. Given a quadruple

$$(\mathbf{A}, \lambda, \nu, t_\alpha) \in \mathfrak{M}(\mathbb{F}_p^{\text{alg}}),$$

fix, as in the proof of Proposition 4.1.1, an  $\mathcal{O}_B$ -linear isogeny  $\rho: A^* \rightarrow A$  for which  $\rho^\vee \circ \lambda \circ \rho$  is an integer multiple of  $\lambda^*$ . Let

$$((\mathfrak{G}, \rho), g) \in \mathfrak{X}(\mathbb{F}_p^{\text{alg}}) \times \overline{G}(\mathbb{A}_f^p)$$

be the pair associated to  $\rho$  and  $(\mathbf{A}, \lambda, \nu)$  as in the proof of Proposition 4.1.1, and define

$$\tau = \rho^{-1} \circ t_\alpha \circ \rho \in \text{End}^0(\mathbf{A}^*).$$

Then  $\tau$  is trace zero and satisfies  $\overline{Q}(\tau) = -\alpha$ , and the condition that  $t_\alpha$  be integral away from  $p$  is equivalent to the condition that  $g \in \Omega(\tau)$ . The isomorphism (32) now takes the quadruple  $(\mathbf{A}, \lambda, \nu, t_\alpha)$  to the point on the right-hand side of (32) corresponding the pair  $((\mathfrak{G}, \rho), g)$  lying in the component of the disjoint union indexed by  $\tau$ .

For every  $\tau_0 \in \overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p$ , viewed as a quasi-endomorphism

$$\tau_0 \in \text{End}(\mathfrak{G}_0^* \times_S S_{/\mathbb{F}_p^{\text{alg}}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p,$$

let  $\mathfrak{h}_m(\tau_0)(S) \subset \mathfrak{h}_m(S)$  be the subset consisting of those pairs  $(\mathfrak{G}_0, \rho_0)$  for which the quasi-endomorphism

$$\rho_0 \circ \tau_0 \circ \rho_0^{-1} \in \text{End}(\mathfrak{G}_0 \times_S S_{/\mathbb{F}_p^{\text{alg}}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$$

lies in the image of the (injective) reduction map

$$\text{End}(\mathfrak{G}_0) \rightarrow \text{End}(\mathfrak{G}_0 \times_S S_{/\mathbb{F}_p^{\text{alg}}}).$$

The set-valued functor  $\mathfrak{h}_m(\tau_0)$  on  $\mathbf{Nilp}_W$  is represented by a closed formal subscheme of  $\mathfrak{h}_m$  (see [KR00, § 2] and [RZ96, ch. 2] for details) and we set

$$\mathfrak{X}_0(\tau_0) = \bigsqcup_{m \in \mathbb{Z}} \mathfrak{h}_m(\tau_0).$$

For each  $\tau \in \overline{V}$ , let  $(\tau_1, \tau_2)$  be the image of  $\tau$  under the isomorphism

$$\overline{V} \otimes_{\mathbb{Q}} \mathbb{Q}_p \cong (\overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p) \times (\overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p). \tag{33}$$

Define closed formal subschemes of  $\mathfrak{X}$  by

$$\mathfrak{X}^1(\tau) = \bigsqcup_{m \in \mathbb{Z}} (\mathfrak{h}_m(\tau_1) \times_W \mathfrak{h}_m), \quad \mathfrak{X}^2(\tau) = \bigsqcup_{m \in \mathbb{Z}} (\mathfrak{h}_m \times_W \mathfrak{h}_m(\tau_2)).$$

As with the isomorphism of functors (32), the functor  $\mathfrak{M}^k$  is represented by the closed formal subscheme of  $\mathfrak{M}$ ,

$$\mathfrak{M}^k \cong \overline{G}(\mathbb{Q}) \backslash \bigsqcup_{\substack{\tau \in \overline{V} \\ \overline{Q}(\tau) = \alpha}} (\mathfrak{X}^k(\tau) \times \Omega(\tau) / \overline{U}^p). \tag{34}$$

If we now set  $\mathfrak{X}(\tau) = \mathfrak{X}(\tau_1) \times_{\mathfrak{X}} \mathfrak{X}(\tau_2)$  then there is a diagram

$$\begin{array}{ccc}
 & \mathfrak{X}^1(\tau) & \\
 & \nearrow & \searrow \\
 \mathfrak{X}(\tau) & & \mathfrak{X} \longrightarrow \mathrm{Spf}(W) \\
 & \searrow & \nearrow \\
 & \mathfrak{X}^2(\tau) & 
 \end{array}$$

in which the square is cartesian, and an isomorphism of functors

$$\widehat{Y}_W \cong \overline{G}(\mathbb{Q}) \backslash \bigsqcup_{\substack{\tau \in \overline{V} \\ \overline{Q}(\tau) = \alpha}} (\mathfrak{X}(\tau) \times \Omega(\tau) / \overline{U}^p).$$

LEMMA 4.1.3. For any  $x \in \mathfrak{M}^k$  the local ring  $\mathcal{O}_{\mathfrak{M},x}$  is regular and

$$\dim \mathcal{O}_{\mathfrak{M}^k,x} < \dim \mathcal{O}_{\mathfrak{M},x}. \tag{35}$$

*Proof.* The regularity of the formal scheme  $\mathfrak{X}$  is clear from its construction (and from the construction of  $\mathfrak{h}_m$  as in [BC92, ch. I] or [KR00, § 1]), and so the regularity of  $\mathfrak{M}$  follows from the uniformization (32). Using the uniformizations (32) and (34), to prove the inequality (35) it suffices to prove

$$\dim \mathcal{O}_{\mathfrak{X}^k(\tau),x} < \dim \mathcal{O}_{\mathfrak{X},x}$$

for every closed point  $x \in \mathfrak{X}^k(\tau)$ . For such an  $x$ ,  $\dim \mathcal{O}_{\mathfrak{X},x} = 3$  while the explicit local equations for the closed formal subscheme  $\mathfrak{h}_m(\tau_k) \hookrightarrow \mathfrak{h}_m$  computed by Kudla and Rapoport [KR00, § 3] (and, for  $p = 2$ , by Kudla, Rapoport and Yang in the appendix to [KRY04, § 11]) show that  $\mathcal{O}_{\mathfrak{X}^k(\tau),x}$  has dimension at most two.  $\square$

### 4.2 The construction of $\mathcal{C}_p$

Continue with the notation of the previous subsection. We have defined formal schemes  $\mathfrak{M}$  and  $\mathfrak{M}^k$  over  $W$  which represent particular functors on  $\mathbf{Nilp}_W$ . The definitions of these functors extend verbatim to functors on  $\mathbf{Nilp}_{\mathbb{Z}_p}$ , and presumably these functors are represented by formal schemes over  $\mathbb{Z}_p$ . However, we will avoid this issue by keeping track of additional descent datum. Let  $\mathrm{Frob} : W \rightarrow W$  be the continuous  $\mathbb{Z}_p$ -algebra isomorphism whose reduction modulo  $pW$  is the absolute Frobenius  $x \mapsto x^p$ . Define a formal  $W$ -scheme  $\mathfrak{M}^{\mathrm{Frob}}$  as the pullback of  $\mathfrak{M}$  by  $\mathrm{Frob}$ , so that

$$\begin{array}{ccc}
 \mathfrak{M}^{\mathrm{Frob}} & \xrightarrow{\mathrm{Frob}} & \mathfrak{M} \\
 \downarrow & & \downarrow \\
 \mathrm{Spf}(W) & \xrightarrow{\mathrm{Frob}} & \mathrm{Spf}(W)
 \end{array}$$

is cartesian. If for any  $r \in \mathbb{Z}^+$  we set  $W_r = W/p^rW$  then the  $W_r$ -scheme  $\mathfrak{M}/W_r$  carries a universal quadruple  $(\mathbf{A}, \lambda, \nu, t_\alpha)$  which we may pull back by  $\mathrm{Frob}$  to yield a quadruple  $(\mathbf{A}^{\mathrm{Frob}}, \lambda^{\mathrm{Frob}}, \nu^{\mathrm{Frob}}, t_\alpha^{\mathrm{Frob}})$  over the scheme  $\mathfrak{M}^{\mathrm{Frob}}/W_r$ . By the universal property of  $\mathfrak{M}$  this quadruple determines a morphism of  $W_r$ -schemes  $\mathfrak{M}^{\mathrm{Frob}}/W_r \rightarrow \mathfrak{M}/W_r$  which is easily seen to be an isomorphism. Letting  $r \rightarrow \infty$  we obtain a canonical isomorphism of formal  $W$ -schemes  $\mathfrak{M}^{\mathrm{Frob}} \cong \mathfrak{M}$ . If  $\mathfrak{F}$  is a coherent  $\mathcal{O}_{\mathfrak{M}}$ -module we say that  $\mathfrak{F}$  is  $\mathrm{Frob}$ -invariant if this isomorphism identifies the pullback

Frob\* $\mathfrak{F}$  with  $\mathfrak{F}$ . Similarly, we may speak of Frob-invariant coherent sheaves on  $\mathfrak{M}^k$  or  $Y/W$ , and Grothendieck’s theory of faithfully flat descent [BLR90, ch. 6] implies that there is a canonical bijection between Frob-invariant coherent  $\mathcal{O}_{Y/W}$ -modules and coherent  $\mathcal{O}_Y$ -modules.

Let  $\mathfrak{F}^k$  be a Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}^k}$ -module, which we view also as a Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}}$ -module annihilated by the ideal sheaf of the closed formal subscheme  $\mathfrak{M}^k$ . It follows from (31) that the Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}}$ -module  $\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}}}(\mathfrak{F}^1, \mathfrak{F}^2)$  on  $\mathfrak{M}$  is annihilated by the ideal sheaf of the formal subscheme  $\widehat{Y}_W \rightarrow \mathfrak{M}$  for every  $\ell \geq 0$ . Thus we may view  $\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}}}(\mathfrak{F}^1, \mathfrak{F}^2)$  as a Frob-invariant coherent  $\mathcal{O}_{\widehat{Y}_W}$ -module and also, by formal GAGA [FGIKNV05, Theorem 8.4.2] and faithfully flat descent, as a coherent  $\mathcal{O}_Y$ -module. Thus we may define

$$[\mathfrak{F}^1 \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathfrak{F}^2] \stackrel{\text{def}}{=} \sum_{\ell \geq 0} (-1)^\ell [\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}}}(\mathfrak{F}^1, \mathfrak{F}^2)] \in \mathbf{K}_0(Y). \tag{36}$$

Our next goal is to prove that the class (36) is supported in dimension one, in the following sense.

DEFINITION 4.2.1. Let  $S$  be a Noetherian scheme and  $[\mathcal{F}] \in \mathbf{K}_0(S)$ . We will say that  $[\mathcal{F}]$  is supported in dimension  $m$  if  $[\mathcal{F}]$  lies in the kernel of the localization map

$$\mathbf{K}_0(S) \rightarrow \mathbf{K}_0(\text{Spec}(\mathcal{O}_{S,\eta}))$$

defined by  $[\mathcal{F}] \mapsto [\mathcal{F}_\eta]$  for every point  $\eta \in S$  with  $\dim \overline{\{\eta\}} > m$ .

LEMMA 4.2.2. Suppose  $D$  is a local Noetherian domain and  $N_1$  and  $N_2$  are finitely generated torsion  $D$ -modules with  $N_2$  of finite length. Then

$$\sum_{\ell \geq 0} (-1)^\ell \text{length}_D \text{Tor}_\ell^D(N_1, N_2) = 0.$$

Proof. The proof is based on that of [KR00, Lemma 4.1]. Let

$$\dots \rightarrow D^{n_1} \rightarrow D^{n_0} \rightarrow N_1 \rightarrow 0$$

be a finite resolution of  $N_1$  by free  $D$ -modules of finite rank. As  $N_1$  is torsion we must have  $\sum (-1)^i n_i = 0$ . As  $N_2$  has finite length, the alternating sum of the lengths of the homology modules of

$$\dots \rightarrow D^{n_1} \otimes_D N_2 \rightarrow D^{n_0} \otimes_D N_2 \rightarrow 0$$

is equal to

$$\sum_{i \geq 0} (-1)^i \cdot \text{length}_D(D^{n_i} \otimes_D N_2) = \text{length}_D(N_2) \cdot \sum_{i \geq 0} (-1)^i n_i = 0. \quad \square$$

PROPOSITION 4.2.3. For  $k \in \{1, 2\}$  let  $\mathfrak{F}^k$  be a Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}^k}$ -module. The class (36) is supported in dimension one.

Proof. Let  $\kappa \in Y$  be the generic point of any two-dimensional component, so that  $\mathcal{O}_{Y,\kappa}$  is Artinian and  $\kappa \in Y/\mathbb{F}_p$ . Choose a point  $\eta \in Y/W$  above  $\kappa$  and consider the commutative diagram

$$\begin{array}{ccccc} \mathbf{K}_0(Y) & \longrightarrow & \mathbf{K}_0(\text{Spec}(\mathcal{O}_{Y,\kappa})) & \longrightarrow & \mathbb{Z} \\ \downarrow & & \downarrow & & \downarrow \text{id} \\ \mathbf{K}_0(Y/W) & \longrightarrow & \mathbf{K}_0(\text{Spec}(\mathcal{O}_{Y/W,\eta})) & \longrightarrow & \mathbb{Z} \end{array}$$

in which the arrows are as follows: the left and middle vertical arrows are flat pullback of coherent sheaves, the upper and lower horizontal arrows on the left are localization, the horizontal arrow on the upper right is the isomorphism  $[\mathcal{F}] \mapsto \text{length}_{\mathcal{O}_{Y, \kappa}}(\mathcal{F})$ , and the horizontal arrow on the lower right is defined similarly. Thus, to check that a class  $[\mathcal{F}] \in \mathbf{K}_0(Y)$  is supported in dimension one, it suffices to check that  $[\mathcal{F}]$  satisfies

$$\text{length}_{\mathcal{O}_{Y/W, \eta}}(\mathcal{F}_{/W, \eta}) = 0$$

for every two-dimensional component  $\overline{\{\eta\}} \in Y/W$ . As the residue field of  $\eta$  has characteristic  $p$  we may identify  $\eta$  with a point of the formal scheme  $\widehat{Y}_W$ . Lemma 4.1.3 implies that  $\mathcal{O}_{\mathfrak{M}, \eta}$  is regular of dimension one, and so is a discrete valuation ring; the same lemma implies that  $\mathcal{O}_{\mathfrak{M}^k, \eta}$ , and hence also  $\mathfrak{F}_\eta^k$ , is a finite length  $\mathcal{O}_{\mathfrak{M}, \eta}$ -module. Applying Lemma 4.2.2 we find

$$\sum_{\ell \geq 0} (-1)^\ell \text{length}_{\mathcal{O}_{\mathfrak{M}, \eta}}(\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}, \eta}}(\mathfrak{F}_\eta^1, \mathfrak{F}_\eta^2)) = 0.$$

This implies  $\text{length}_{\mathcal{O}_{Y/W, \eta}}(\mathfrak{F}^1 \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathfrak{F}^2) = 0$ , completing the proof. □

LEMMA 4.2.4. *Suppose  $[\mathcal{F}] \in \mathbf{K}_0(Y)$  is supported in dimension one. Then the image of  $[\mathcal{F}]$  under the homomorphism*

$$R\phi_* : \mathbf{K}_0(Y) \rightarrow \mathbf{K}_0^Y(M)$$

*of (8) lies in  $F^2\mathbf{K}_0^Y(M)$ .*

*Proof.* Lemma 2.2.2 implies that  $[\mathcal{F}]$  lies in the image of  $\mathbf{K}_0(Z) \rightarrow \mathbf{K}_0(Y)$  for some closed subscheme  $Z \rightarrow Y$  of dimension one. But then  $R\phi_*[\mathcal{F}]$  lies in the image of

$$\mathbf{K}_0^Z(M) = F^2\mathbf{K}_0^Z(M) \rightarrow F^2\mathbf{K}_0^Y(M). \quad \square$$

We now have a machine for producing elements of  $\text{CH}_Y^2(M)$ . If  $[\mathcal{F}] \in \mathbf{K}_0(Y)$  is supported in dimension one we define, using Lemma 4.2.4,

$$\text{cl}(\mathcal{F}) \in \text{CH}_Y^2(M)$$

to be the image of  $R\phi_*[\mathcal{F}]$  under the map  $F^2\mathbf{K}_0^Y(M) \rightarrow \text{CH}_Y^2(M)$  induced by the isomorphism (9). In particular, if  $\mathfrak{F}^k$  is a Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}^k}$ -module for  $k \in \{1, 2\}$  then we may form, using Proposition 4.2.3,

$$\text{cl}(\mathfrak{F}^1 \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathfrak{F}^2) \in \text{CH}_Y^2(M). \quad (37)$$

The issue is which coherent sheaves to choose, and we make the obvious choice  $\mathfrak{F}^k = \mathcal{O}_{\mathfrak{M}^k}$ . Applying the construction (37), define

$$C_p = \text{cl}(\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}) \in \text{CH}_Y^2(M).$$

The group  $H = U^{\max}/U$  acts on  $\mathfrak{M}$  and  $\mathfrak{M}^k$  by permuting  $U$ -level structures, and the cycle class  $C_p$  is  $H$ -invariant by construction. By Lemma 2.1.2,  $C_p$  arises as the flat pullback of a cycle class

$$C_p \in \text{CH}_{Y/\mathbb{Z}_p}^2(\mathcal{M}/\mathbb{Z}_p). \quad (38)$$

PROPOSITION 4.2.5. *The homomorphism (12) takes  $C_p$  to the cycle  $\mathcal{C}_{\mathbb{Q}} \times_{\mathbb{Q}} \mathbb{Q}_p$  constructed from (22).*

*Proof.* First suppose that we have a closed point  $x_0 \in \mathfrak{X}_0$  and a  $\tau_0 \in \overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p$ . In the language of [KR00] the point  $x_0$  is either *ordinary* or *superspecial*. Let  $R_0$  denote the completion of the local

ring  $\mathcal{O}_{\mathfrak{X}_0, x_0}$  under the topology induced by its maximal ideal, and let  $N_0$  denote the completion of the stalk  $\mathcal{O}_{\mathfrak{X}_0(\tau_0), x_0}$ . Let  $P_0$  denote the quotient of  $N_0$  by its maximal  $W$ -torsion submodule. Kudla and Rapoport [KR00, §3] and Kudla, Rapoport and Yang (in the case  $p = 2$ ; see the appendix to [KRY04, §11]) determine the structure of  $N_0$ , and hence also  $P_0$ , quite explicitly. If  $x_0$  is ordinary then  $R_0 \cong W[[s]]$  and either  $P_0 = 0$  or  $P_0 \cong W[[s]]/(s)$ . If  $x_0$  is superspecial then  $R_0 \cong W[[s, t]]/(st - p)$  and either  $P_0 = 0$  or

$$P_0 \cong W[[s, t]]/(st - p, s + ap + ut)$$

for some  $a, u \in W$  with  $u$  a unit.

Now fix a closed point  $x \in \mathfrak{X}$  contained in both  $\mathfrak{X}^1(\tau)$  and  $\mathfrak{X}^2(\tau)$ , let  $R$  be the completed local ring of  $\mathcal{O}_{\mathfrak{X}}$  at  $x$ , let  $N^k$  be the completed stalk of  $\mathcal{O}_{\mathfrak{X}^k(\tau)}$  at  $x$ , and let  $P^k$  be the quotient of  $N^k$  by its maximal  $W$ -torsion submodule. The geometric point

$$\mathrm{Spf}(k(x)) \rightarrow \mathfrak{X} \rightarrow \mathfrak{X}_0 \times_W \mathfrak{X}_0$$

determines an ordered pair of points  $(x_1, x_2)$  of  $\mathfrak{X}_0$ , and we consider separately the four possibilities for the pair of points appearing. First suppose both points are ordinary. Then

$$R \cong W[[s_1, s_2]], \quad P^1 \cong R/(s_1), \quad P^2 \cong R/(s_2).$$

Using the projective resolution

$$0 \rightarrow R \xrightarrow{s_1} R \rightarrow P_1 \rightarrow 0$$

of  $P_1$  we find that the  $R$ -modules  $\mathrm{Tor}_\ell^R(P^1, P^2)$  are equal to the homology modules of the complex

$$0 \rightarrow R/(s_2) \xrightarrow{s_1} R/(s_2) \rightarrow 0$$

and therefore

$$\mathrm{Tor}_\ell^R(P^1, P^2) = 0 \quad \forall \ell > 0. \tag{39}$$

Next suppose that  $x_1$  is ordinary and  $x_2$  is superspecial. Then

$$R \cong W[[s_1, s_2, t_2]]/(s_2t_2 - p), \quad P^1 \cong R/(s_1), \quad P^2 \cong R/(s_2 + a_2p + u_2t_2)$$

and using the same projective resolution of  $P_1$  as in the ordinary/ordinary case we find that  $\mathrm{Tor}_\ell^R(P^1, P^2)$  is given by the homology of

$$0 \rightarrow R/(s_2 + a_2 + u_2t_2) \xrightarrow{s_1} R/(s_2 + a_2p + u_2t_2) \rightarrow 0.$$

Again we see that (39) holds. Obviously the case of  $x_1$  superspecial and  $x_2$  ordinary is similar. Finally suppose that  $x_1$  and  $x_2$  are both superspecial. Then

$$R \cong W[[s_1, t_1, s_2, t_2]]/(s_1t_1 - p, s_2t_2 - p), \quad P^k \cong R/(s_k + a_kp + u_kt_k)$$

for some  $a_k \in W^\times$ . Using the projective resolution

$$0 \rightarrow R \xrightarrow{s_1 + a_1p + u_1t_1} R \rightarrow P^1 \rightarrow 0$$

of  $P^1$  one again verifies (39). We deduce that, in all cases, whenever  $\ell > 0$  the  $R$ -module  $\mathrm{Tor}_\ell^R(N^1, N^2)$  is  $W$ -torsion. Using the  $p$ -adic uniformizations (32) and (34) it follows that

$$\mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}^x}}(\mathcal{O}_{\mathfrak{M}^1, x}, \mathcal{O}_{\mathfrak{M}^2, x})$$

is  $W$ -torsion for all  $\ell > 0$  and all closed points  $x \in \mathfrak{M}$ .

Fix  $\ell > 0$  and abbreviate

$$\mathcal{T} = \mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{O}_{\mathfrak{M}^1}, \mathcal{O}_{\mathfrak{M}^2})$$



viewed either as a coherent  $\mathcal{O}_Y$ -module or a coherent  $\mathcal{O}_{\widehat{Y}/W}$ -module. A closed point  $y \in Y$  is contained in the special fiber of  $Y$ , and so may be viewed as a point of  $\widehat{Y}$ . For any point  $x \in \widehat{Y}/W$  above  $y$ , the previous paragraph shows that  $\mathcal{T}_x$  is  $W$ -torsion, and it follows that  $\mathcal{T}_y$  is  $\mathbb{Z}_p$ -torsion. As this holds for all closed points  $y \in Y$ , it holds for *all* points  $y \in Y$ , including those contained in the generic fiber. But if  $y \in Y$  lies in the generic fiber then  $\mathcal{T}_y$  is a  $\mathbb{Q}_p$ -vector space, and hence  $\mathcal{T}_y = 0$ . Thus the pullback  $\mathcal{T}/\mathbb{Q}_p$  of  $\mathcal{T}$  by  $Y/\mathbb{Q}_p \rightarrow Y$  has trivial stalks, and so is trivial. In particular,  $[\mathcal{T}]$  lies in the kernel of the flat pullback  $\mathbf{K}_0(Y) \rightarrow \mathbf{K}_0(Y/\mathbb{Q}_p)$  and so the class

$$[\mathcal{O}_Y] = [\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}} \mathcal{O}_{\mathfrak{M}^2}] = [\mathrm{Tor}_0^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{O}_{\mathfrak{M}^1}, \mathcal{O}_{\mathfrak{M}^2})]$$

has the same image as  $[\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}]$  under  $\mathbf{K}_0(Y) \rightarrow \mathbf{K}_0(Y/\mathbb{Q}_p)$ . The claim now follows exactly as in the proof of Lemma 3.3.2.  $\square$

### 4.3 Computing the pullback

We assume throughout § 4.3 that  $Y_0/\mathbb{Q}_p = \emptyset$ . Continuing with the notation of the previous subsections, our next task is to compute  $\mathrm{deg}_p(i^*C_p)$  where  $i^*$  is the pullback of (10) and  $\mathrm{deg}_p$  is the composition (29). Let  $\mu : Y_0 \rightarrow \mathrm{Spec}(\mathbb{Z}_p)$  be the structure map. The assumption that  $Y_0/\mathbb{Q}_p = \emptyset$  implies that for any coherent  $\mathcal{O}_{Y_0}$ -module  $\mathcal{F}$  the higher direct image  $R^\ell \mu_* \mathcal{F}$  is a torsion  $\mathbb{Z}_p$ -module, and so there is a push-forward

$$R\mu_* : \mathbf{K}_0(Y_0) \rightarrow \mathbf{K}_0^{\{p\}}(\mathrm{Spec}(\mathbb{Z}_p))$$

defined by

$$R\mu_*[\mathcal{F}] = \sum_{\ell \geq 0} (-1)^\ell [R^\ell \mu_* \mathcal{F}].$$

The composition of the push-forward with the isomorphism  $\mathbf{K}_0^{\{p\}}(\mathrm{Spec}(\mathbb{Z}_p)) \cong \mathbb{Z}$  defined by  $[\mathcal{G}] \mapsto \mathrm{length}_{\mathbb{Z}_p}(\mathcal{G})$  defines the *Euler characteristic*  $\chi : \mathbf{K}_0(Y_0) \rightarrow \mathbb{Z}$ ,

$$\chi(\mathcal{F}) = \sum_{\ell \geq 0} (-1)^\ell \mathrm{length}_{\mathbb{Z}_p}(R^\ell \mu_* \mathcal{F}).$$

Let  $\widehat{M}_0$  and  $\widehat{Y}_0$  denote the formal completions of  $M_0$  and  $Y_0$  along their special fibers, respectively. Define formal  $W$ -schemes ( $k \in \{1, 2\}$ )

$$\mathfrak{M}_0 = \widehat{M}_0/W \times_{\widehat{M}/W} \mathfrak{M}, \quad \mathfrak{M}_0^k = \widehat{M}_0/W \times_{\widehat{M}/W} \mathfrak{M}^k,$$

so that there is an isomorphism

$$\widehat{Y}_0/W \cong \mathfrak{M}_0 \times_{\mathfrak{M}} \widehat{Y}/W$$

and a diagram

$$\begin{array}{ccccc}
 & & \mathfrak{M}_0^1 & & \\
 & \nearrow & & \searrow & \\
 \widehat{Y}_0/W & & & & \mathfrak{M}_0 \longrightarrow \widehat{M}_0/W \\
 & \searrow & & \nearrow & \\
 & & \mathfrak{M}_0^2 & & 
 \end{array}$$

in which the square is cartesian. For each  $(s_1, s_2) \in \bar{V}_0$  we view  $\tau = s_1\varpi_1 + s_2\varpi_2$  as an element of  $\bar{V} \cong \bar{V}_0 \otimes_{\mathbb{Q}} F$  and let  $(\tau_1, \tau_2) \in (\bar{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p)^2$  denote the image of  $\tau$  under (33). If we define

$$\Omega_0(s_1, s_2) = \Omega(\tau) \cap \bar{G}_0(\mathbb{A}_F^p)$$

then, as in the discussion surrounding [KR00, (8.17)], there is a Cerednik–Drinfeld style uniformization

$$\mathfrak{M}_0 \cong \bar{G}_0(\mathbb{Q}) \backslash \bigsqcup_{\substack{(s_1, s_2) \in \bar{V}_0 \\ \bar{Q}(\tau) = \alpha}} (\mathfrak{X}_0 \times \Omega_0(s_1, s_2) \bar{U}^{\max, p} / \bar{U}^p). \tag{40}$$

One obtains a similar uniformization of  $\mathfrak{M}_0^k$  by replacing  $\mathfrak{X}_0$  with  $\mathfrak{X}_0(\tau_k)$ , and a uniformization of  $\widehat{Y}_{0/W}$  by replacing  $\mathfrak{X}_0$  with  $\mathfrak{X}_0(\tau_1) \times_{\mathfrak{X}_0} \mathfrak{X}_0(\tau_2)$ .

Define a homomorphism  $\mathbf{K}_0(Y) \rightarrow \mathbf{K}_0(Y_0)$  as follows. If  $\mathcal{F}$  is a coherent  $\mathcal{O}_Y$ -module then, by pulling back to  $Y/W$  and passing to the formal completion, we may view  $\mathcal{F}$  as a  $\sigma$ -invariant coherent  $\mathcal{O}_{\mathfrak{M}}$ -module annihilated by the ideal sheaf of the closed formal subscheme  $\widehat{Y}_{0/W} \rightarrow \mathfrak{M}$ . For each  $\ell \geq 0$  the coherent  $\mathcal{O}_{\mathfrak{M}}$ -module  $\text{Tor}_{\ell}^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{F}, \mathcal{O}_{\mathfrak{M}_0})$  is Frob-invariant and is annihilated by the ideal sheaf of the closed formal subscheme  $\widehat{Y}_{0/W} \rightarrow \mathfrak{M}$ , and so by formal GAGA and faithfully flat descent may be viewed as a coherent  $\mathcal{O}_{Y_0}$ -module. Thus we may define

$$[\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}] \stackrel{\text{def}}{=} \sum_{\ell \geq 0} (-1)^{\ell} [\text{Tor}_{\ell}^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{F}, \mathcal{O}_{\mathfrak{M}_0})] \in \mathbf{K}_0(Y_0),$$

and one checks that  $[\mathcal{F}] \mapsto [\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]$  defines a map  $\mathbf{K}_0(Y) \rightarrow \mathbf{K}_0(Y_0)$ .

PROPOSITION 4.3.1. *If  $[\mathcal{F}] \in \mathbf{K}_0(Y)$  is supported in dimension one then the class  $[\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]$  is supported in dimension zero and*

$$\text{deg}_p(i^* \text{cl}(\mathcal{F})) = \chi(\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}) \tag{41}$$

where  $i^*$  is the pullback of (10) and  $\text{deg}_p$  is the composition (29).

*Proof.* Using Lemma 2.2.2 we are reduced to the case where  $\mathcal{F}$  is supported in dimension one in the usual sheaf-theoretic sense. To show that  $[\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]$  is supported in dimension zero we must show that it lies in the kernel of

$$\mathbf{K}_0(Y_0) \rightarrow \mathbf{K}_0(\text{Spec}(\mathcal{O}_{Y_0, \kappa})) \cong \mathbb{Z}$$

for every  $\kappa \in Y_0$  satisfying  $\dim \overline{\{\kappa\}} = 1$ , where the isomorphism is

$$[\mathcal{G}] \mapsto \text{length}_{\mathcal{O}_{Y_0, \eta}}(\mathcal{G}).$$

As we assume that  $Y_{0/\mathbb{Q}_p} = \emptyset$ , such a  $\kappa$  must lie in the special fiber of  $Y_0$ , and thus we may view  $\kappa$  as a point of the formal completion  $\widehat{Y}_0$ . Given a point  $\eta \in \widehat{Y}_{0/W}$  lying above  $\kappa$  the hypothesis that  $\mathcal{F}$  is supported in dimension one implies that  $\mathcal{F}_{/W, \eta}$  is a finite length  $\mathcal{O}_{\widehat{Y}_{0/W, \eta}}$ -module, and Lemma 4.2.2 then shows that

$$\sum_{\ell \geq 0} (-1)^{\ell} \text{length}_{\mathcal{O}_{\mathfrak{M}, \eta}}(\text{Tor}_{\ell}^{\mathcal{O}_{\mathfrak{M}, \eta}}(\mathcal{F}_{/W, \eta}, \mathcal{O}_{\mathfrak{M}_0, \eta})) = 0.$$

Exactly as in the proof of Proposition 4.2.3, it follows that

$$\text{length}_{\mathcal{O}_{Y_0, \kappa}}([\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]) = 0.$$

This completes the proof of the first claim.

It remains to prove (41). Let  $\tilde{\mathbf{K}}_0(Y) \subset \mathbf{K}_0(Y)$  be the subgroup of classes supported in dimension one, and similarly let  $\tilde{\mathbf{K}}_0(Y_0) \subset \mathbf{K}_0(Y_0)$  be the subgroup of classes supported in dimension zero. There is a commutative diagram

$$\begin{CD} \tilde{\mathbf{K}}_0(Y) @>>> \tilde{\mathbf{K}}_0(Y_0) \\ @V R\phi_* VV @VV R\phi_{0*} V \\ F^2\mathbf{K}_0^Y(M) @>i^*>> F^2\mathbf{K}_0^{Y_0}(M_0) \\ @VVV @VVV \\ \text{CH}_Y^2(M) @>i^*>> \text{CH}_{Y_0}^2(M_0) \end{CD}$$

in which the top horizontal arrow is  $[\mathcal{F}] \mapsto [\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]$  and the middle horizontal arrow is the pullback of § 2.2, given by

$$[\mathcal{G}] \mapsto [\mathcal{G} \otimes_{\mathcal{O}_M}^L \mathcal{O}_{M_0}] = \sum_{\ell \geq 0} (-1)^\ell [\text{Tor}_\ell^{\mathcal{O}_M}(\mathcal{G}, \mathcal{O}_{M_0})].$$

Thus the left-hand side of (41) is equal to the image of  $[\mathcal{F} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}]$  under the composition

$$\tilde{\mathbf{K}}_0(Y_0) \xrightarrow{R\phi_{0*}} F^2\mathbf{K}_0^{Y_0}(M_0) \rightarrow \text{CH}_{Y_0}^2(M_0) \xrightarrow{\text{deg}_p} \mathbb{Q}.$$

Using Lemma 2.2.2 to reduce to the case in which  $\mathcal{G}$  is a skyscraper sheaf supported at a closed point of  $Y_0$ , one checks that this composition is simply  $[\mathcal{G}] \mapsto \chi(\mathcal{G})$ , completing the proof.  $\square$

Given Frob-invariant coherent  $\mathcal{O}_{\mathfrak{M}_0^k}$ -modules  $\mathfrak{F}_0^k$  for  $k \in \{1, 2\}$ , define

$$[\mathfrak{F}_0^1 \otimes_{\mathcal{O}_{\mathfrak{M}_0}}^L \mathfrak{F}_0^2] = \sum_{\ell \geq 0} (-1)^\ell [\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}_0}}(\mathfrak{F}_0^1, \mathfrak{F}_0^2)] \in \mathbf{K}_0(Y_0)$$

as in the construction (36). This class is supported in dimension zero, by imitating the proof of Proposition 4.2.3.

LEMMA 4.3.2. *We have the equality of Euler characteristics*

$$\chi((\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}) \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}) = \chi(\mathcal{O}_{\mathfrak{M}_0^1} \otimes_{\mathcal{O}_{\mathfrak{M}_0}}^L \mathcal{O}_{\mathfrak{M}_0^2}).$$

*Proof.* Elementary homological algebra shows that

$$\begin{aligned} &(\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}) \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0} \\ &= (\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}) \otimes_{\mathcal{O}_{\mathfrak{M}}}^L (\mathcal{O}_{\mathfrak{M}^2} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}) \end{aligned}$$

in the derived category of locally free quasi-coherent  $\mathcal{O}_{\mathfrak{M}}$ -modules. From this it follows that

$$\begin{aligned} &\chi((\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}) \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}_0}) \\ &= \sum_{i,j,\ell} (-1)^{i+j+\ell} \chi(\text{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}_0}}(\text{Tor}_i^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{O}_{\mathfrak{M}^1}, \mathcal{O}_{\mathfrak{M}_0}), \text{Tor}_j^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{O}_{\mathfrak{M}^2}, \mathcal{O}_{\mathfrak{M}_0}))). \end{aligned}$$

We claim that

$$\text{Tor}_i^{\mathcal{O}_{\mathfrak{M}}}(\mathcal{O}_{\mathfrak{M}^1}, \mathcal{O}_{\mathfrak{M}_0}) \cong \begin{cases} \mathcal{O}_{\mathfrak{M}_0^1} & \text{if } i = 0, \\ 0 & \text{otherwise,} \end{cases}$$

for all  $i \geq 0$  (and similarly for  $\mathcal{O}_{\mathfrak{M}^2}$ ). When  $i = 0$  this is clear from  $\mathfrak{M}^1 \times_{\mathfrak{M}} \mathfrak{M}_0 \cong \mathfrak{M}_0^1$ . When  $i > 0$  it suffices to check the vanishing at stalks of the sheaf on the left. Using the uniformization (32)

we are then reduced to proving the vanishing of

$$\mathrm{Tor}_i^{\mathcal{O}_{\mathfrak{h}_m \times_W \mathfrak{h}_m}}(\mathcal{O}_{\mathfrak{h}_m(\tau_1) \times_W \mathfrak{h}_m}, \mathcal{O}_{\mathfrak{h}_m}) = 0 \tag{42}$$

for every  $\tau_1 \in \overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p$ . Here we view  $\mathfrak{h}_m$  as a closed formal subscheme of  $\mathfrak{h}_m \times_W \mathfrak{h}_m$  using the diagonal embedding, and hence view  $\mathcal{O}_{\mathfrak{h}_m}$  as a coherent  $\mathcal{O}_{\mathfrak{h}_m \times_W \mathfrak{h}_m}$ -module. The key observation is that the flatness of  $\mathfrak{h}_m \rightarrow \mathrm{Spf}(W)$  implies the flatness of the first projection  $\pi_1 : \mathfrak{h}_m \times_W \mathfrak{h}_m \rightarrow \mathfrak{h}_m$ . Choosing a resolution

$$\mathcal{P}^\bullet \rightarrow \mathcal{O}_{\mathfrak{h}_m(\tau_1)}$$

of  $\mathcal{O}_{\mathfrak{h}_m(\tau_1)}$  by locally free quasi-coherent  $\mathcal{O}_{\mathfrak{h}_m}$ -modules the pullback complex

$$\pi_1^* \mathcal{P}^\bullet \rightarrow \mathcal{O}_{\mathfrak{h}_m(\tau_1) \times_W \mathfrak{h}_m}$$

is a locally free resolution of  $\mathcal{O}_{\mathfrak{h}_m(\tau_1) \times_W \mathfrak{h}_m}$ . Using this resolution to compute the left-hand side of (42) and using the isomorphism

$$(\pi_1^* \mathcal{P}) \otimes_{\mathcal{O}_{\mathfrak{h}_m \times_W \mathfrak{h}_m}} \mathcal{O}_{\mathfrak{h}_m} \cong \mathcal{P}$$

for any quasi-coherent  $\mathcal{O}_{\mathfrak{h}_m}$ -module  $\mathcal{P}$ , (42) follows.

Combining the previous two paragraphs we are now left with

$$\chi((\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^2}) \otimes_{\mathcal{O}_{\mathfrak{M}}}^L \mathcal{O}_{\mathfrak{M}^0}) = \sum_{\ell} (-1)^\ell \chi(\mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}^0}}(\mathcal{O}_{\mathfrak{M}^1}, \mathcal{O}_{\mathfrak{M}^2}))$$

as desired. □

**COROLLARY 4.3.3.** *We have*

$$\mathrm{deg}_p(i^* C_p) = \chi(\mathcal{O}_{\mathfrak{M}^1} \otimes_{\mathcal{O}_{\mathfrak{M}^0}}^L \mathcal{O}_{\mathfrak{M}^2}).$$

*Proof.* Combine Proposition 4.3.1 and Lemma 4.3.2. □

We now turn to the calculation of the right-hand side of the equality of Corollary 4.3.3, which we reduce to calculations of Kudla, Rapoport and Yang. Let  $\mathrm{Tr}$  be the reduced trace on  $\overline{B}_0$ . Define a  $\mathbb{Q}$ -valued bilinear form on  $\overline{V}_0$ ,

$$[s_1, s_2] = -\mathrm{Tr}(s_1 s_2),$$

and for each  $T \in \mathrm{Sym}_2(\mathbb{Z})^\vee$  let  $\overline{V}_0(T) \subset \overline{V}_0 \times \overline{V}_0$  be the set of pairs  $(s_1, s_2)$  such that (20) holds. It follows from [Lam05, Theorem III.3.1] that the group  $\overline{G}_0(\mathbb{Q})$  acts transitively on  $\overline{V}_0(T)$  by conjugation. Define a  $\mathbb{Z}[1/p]$ -lattice  $\overline{L}_0 \subset \overline{V}_0$ ,

$$\overline{L}_0 = \{v \in \overline{V}_0 \mid \widehat{\Lambda}_0^p \cdot \iota_0(v) \subset \widehat{\Lambda}_0^p\},$$

and let  $\overline{L}_0(T) = \overline{V}_0(T) \cap (\overline{L}_0 \times \overline{L}_0)$ . Define also a discrete subgroup,

$$\overline{\Gamma}_0 = \{\gamma \in \overline{G}_0(\mathbb{Q}) \mid \widehat{\Lambda}_0^p \cdot \iota_0(\gamma) = \widehat{\Lambda}_0^p\},$$

and note that  $\overline{\Gamma}_0$  acts on  $\overline{L}_0$  and on  $\overline{L}_0(T)$  by conjugation.

Let  $\overline{Q}_0(\tau_0) = -\tau_0^2$  be the quadratic form on  $\overline{V}_0$  associated to  $[\cdot, \cdot]$ . Suppose that  $T \in \mathrm{Sym}_2(\mathbb{Z})^\vee$  is nonsingular, and that  $s_1, s_2 \in \overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p$  satisfy both  $\overline{Q}_0(s_k) \in \mathbb{Z}_p - \{0\}$  and the relation (20). In particular, this implies that  $s_1$  and  $s_2$  are linearly independent. Letting  $\mu : \mathfrak{h}_m \rightarrow \mathrm{Spf}(W)$  denote the structure map, a theorem of Kudla and Rapoport [KR00, Theorem 6.1] (and using the appendix to [KRY06, ch. 6] for the case  $p = 2$ ) shows that for any  $m \in \mathbb{Z}$  the integer

$$e_p(T) \stackrel{\mathrm{def}}{=} \sum_{k, \ell \geq 0} (-1)^{k+\ell} \mathrm{length}_W R^\ell \mu_* \mathrm{Tor}_k^{\mathcal{O}_{\mathfrak{h}_m}}(\mathcal{O}_{\mathfrak{h}_m(s_1)}, \mathcal{O}_{\mathfrak{h}_m(s_2)})$$

depends only on the isomorphism class of the rank two quadratic space

$$(\mathbb{Z}_p \oplus \mathbb{Z}_p, T) \cong (\mathbb{Z}_p s_1 + \mathbb{Z}_p s_2, \overline{Q}_0)$$

and not on  $s_1, s_2$  themselves (and all  $W$ -modules appearing in the sum have finite length). Furthermore,  $e_p(T)$  is independent of  $m$  by the argument following [KR00, (8.40)]. In the notation of [KRY06, ch. 6.2] we have  $\nu_p(T) = 2 \cdot e_p(T)$ .

For each  $T \in \Sigma(\alpha)$ , define  $Z(T) = \mathcal{Z}(T)_{/\mathbb{Z}_p} \times_{\mathcal{M}_{/\mathbb{Z}_p}} M$ . The decomposition (21) induces an analogous decomposition of  $Y_0$ , and hence an isomorphism

$$\mathbf{K}_0(Y_0) \cong \bigoplus_{T \in \Sigma(\alpha)} \mathbf{K}_0(Z(T)). \tag{43}$$

For any  $[\mathcal{G}] \in \mathbf{K}_0(Y_0)$ , let  $\chi_T(\mathcal{G})$  denote the Euler characteristic of the projection of  $[\mathcal{G}]$  to the summand  $\mathbf{K}_0(Z(T))$ .

PROPOSITION 4.3.4. *Assume that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic, so that  $Y_{0/\mathbb{Q}_p} = \emptyset$  and  $\Sigma(\alpha)$  contains no singular matrices by Lemma 3.1.5. If  $T \in \Sigma(\alpha)$  satisfies (20) for some  $s_1, s_2 \in \overline{V}_0$  then*

$$\frac{1}{|H|} \cdot \chi_T(\mathcal{O}_{\mathfrak{M}_0^1} \otimes_{\mathcal{O}_{\mathfrak{M}_0}}^L \mathcal{O}_{\mathfrak{M}_0^2}) = e_p(T) \cdot |\overline{\Gamma}_0 \backslash \overline{L}_0(T)|.$$

*Proof.* The proof is based on that of [KR00, Theorem 8.5]. Abbreviate  $[\mathfrak{F}_T]$  for the projection of  $[\mathcal{O}_{\mathfrak{M}_0^1} \otimes_{\mathcal{O}_{\mathfrak{M}_0}}^L \mathcal{O}_{\mathfrak{M}_0^2}]$  to the summand of (43) indexed by  $T$  and let

$$\mu : Y_0 \rightarrow \text{Spec}(\mathbb{Z}_p)$$

be the structure map so that

$$\chi_T(\mathcal{O}_{\mathfrak{M}_0^1} \otimes_{\mathcal{O}_{\mathfrak{M}_0}}^L \mathcal{O}_{\mathfrak{M}_0^2}) = \text{length}_{\mathbb{Z}_p}(R\mu_* \mathfrak{F}_T).$$

Let  $\widehat{Z}(T)$  denote the formal completion of  $Z(T)$  along its special fiber. After formal completion and base change to  $W$  the morphism  $Z(T) \rightarrow M_0$  admits a factorization

$$\widehat{Z}(T)_{/W} \rightarrow \mathfrak{M}_0(T) \rightarrow \widehat{M}_{0/W}$$

in which the open and closed formal subscheme  $\mathfrak{M}_0(T)$  of  $\mathfrak{M}_0$  is defined, following (40) and [KR00, (8.17)], as

$$\mathfrak{M}_0(T) \cong \overline{G}_0(\mathbb{Q}) \backslash \bigsqcup_{(s_1, s_2) \in \overline{V}_0(T)} (\mathfrak{X}_0 \times \Omega_0(s_1, s_2) \overline{U}^{\max, p} / \overline{U}^p).$$

The closed formal subscheme  $\widehat{Z}(T)_{/W} \rightarrow \mathfrak{M}_0(T)$  is obtained by replacing  $\mathfrak{X}_0$  with

$$\mathfrak{X}_0(\tau_1) \times_{\mathfrak{X}_0} \mathfrak{X}_0(\tau_2)$$

as in the comments following (40), where  $(\tau_1, \tau_2) \in (\overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p)^2$  is the image of  $\tau = s_1 \varpi_1 + s_2 \varpi_2 \in \overline{V}$  under (33). Similarly, we define

$$\mathfrak{M}_0^k(T) \cong \overline{G}_0(\mathbb{Q}) \backslash \bigsqcup_{(s_1, s_2) \in \overline{V}_0(T)} (\mathfrak{X}_0(\tau_k) \times \Omega_0(s_1, s_2) \overline{U}^{\max, p} / \overline{U}^p)$$

so that there is a diagram

$$\begin{array}{ccccc}
 & & \mathfrak{M}_0^1(T) & & \\
 & \nearrow & & \searrow & \\
 \widehat{Z}(T)/W & & & & \mathfrak{M}_0(T) \longrightarrow \mathfrak{M}_0 \\
 & \searrow & & \nearrow & \\
 & & \mathfrak{M}_0^2(T) & & 
 \end{array}$$

in which the square is cartesian. We have the equality in  $\mathbf{K}_0(Z(T))$ ,

$$[\mathfrak{F}_T] = \sum_{\ell \geq 0} (-1)^\ell [\mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{M}_0(T)}}(\mathcal{O}_{\mathfrak{M}_0^1(T)}, \mathcal{O}_{\mathfrak{M}_0^2(T)})].$$

Fix one pair  $(s_1, s_2) \in \overline{V}_0(T)$ . As all such pairs are conjugate by  $\overline{G}_0(\mathbb{Q})$  with stabilizer  $\mathbb{Q}^\times$  (using the linear independence of  $\{s_1, s_2\}$ ), there are isomorphisms

$$\begin{aligned}
 \mathfrak{M}_0(T) &\cong \mathbb{Q}^\times \backslash (\mathfrak{X}_0 \times \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p) \\
 &\cong (\mathfrak{h}_0 \sqcup \mathfrak{h}_1) \times (\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p)
 \end{aligned}$$

where  $\mathbb{Q}^b = \{x \in \mathbb{Q}^\times \mid \mathrm{ord}_p(x) = 0\}$ . For the second isomorphism we have used the fact that the action of  $z \in \mathbb{Q}^\times$  on  $\mathfrak{X}_0$  takes  $\mathfrak{h}_m$  isomorphically to  $\mathfrak{h}_{m-2\mathrm{ord}_p(z)}$ . Under these identifications we have

$$\mathfrak{M}_0^k(T) \cong (\mathfrak{h}_0(\tau_k) \sqcup \mathfrak{h}_1(\tau_k)) \times (\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p)$$

and it follows that  $\mathrm{length}_{\mathbb{Z}_p}(R\mu_*\mathfrak{F}_T)$  is equal to

$$\begin{aligned}
 \mathrm{length}_{\mathbb{Z}_p}(R\mu_*\mathfrak{F}_T) &= |\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p| \\
 &\times \left[ \sum_{k, \ell \geq 0} (-1)^{k+\ell} \mathrm{length}_W R^k \mu_* \mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{h}_0}}(\mathcal{O}_{\mathfrak{h}_0}(\tau_1), \mathcal{O}_{\mathfrak{h}_0}(\tau_2)) \right. \\
 &\quad \left. + \sum_{k, \ell \geq 0} (-1)^{k+\ell} \mathrm{length}_W R^k \mu_* \mathrm{Tor}_\ell^{\mathcal{O}_{\mathfrak{h}_1}}(\mathcal{O}_{\mathfrak{h}_1}(\tau_1), \mathcal{O}_{\mathfrak{h}_1}(\tau_2)) \right] \\
 &= 2 \cdot |\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p| \cdot e_p(T)
 \end{aligned}$$

where we have used the equality  $\mathbb{Z}_p s_1 + \mathbb{Z}_p s_2 = \mathbb{Z}_p \tau_1 + \mathbb{Z}_p \tau_2$  of lattices in  $\overline{V}_0 \otimes_{\mathbb{Q}} \mathbb{Q}_p$  for the final equality. It now suffices to prove

$$2 \cdot |\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p| = |H| \cdot |\overline{\Gamma}_0 \backslash \overline{L}_0(T)|. \tag{44}$$

By strong approximation any element of  $\Omega_0(s_1, s_2) \overline{U}^{\max,p}$  has the form  $\gamma u$  with

$$\gamma \in \overline{G}_0(\mathbb{Q}) \cap \Omega_0(s_1, s_2)$$

and  $u \in \overline{U}^{\max,p}$ . The element  $\gamma$  then satisfies

$$(\gamma^{-1} s_1, \gamma^{-1} s_2) \in \overline{L}_0(T)$$

and we have a well-defined function

$$\mathbb{Q}^b \backslash \Omega_0(s_1, s_2) \overline{U}^{\max,p} / \overline{U}^p \rightarrow \overline{\Gamma}_0 \backslash \overline{L}_0(T)$$

which takes the double coset of  $\gamma u$  to  $\overline{\Gamma}_0 \cdot (\gamma^{-1} s_1, \gamma^{-1} s_2)$ . The surjectivity of this function follows

easily from the transitivity of  $\overline{G}_0(\mathbb{Q})$  on  $\overline{V}_0(T)$ . The fiber over  $\overline{\Gamma}_0 \cdot (\gamma^{-1}s_1, \gamma^{-1}s_2)$  is in bijection with  $\{\pm 1\} \backslash \gamma \overline{U}^{\max,p} / \overline{U}^p$ , and (44) follows.  $\square$

### 5. Pullbacks of arithmetic cycles

#### 5.1 Construction of arithmetic cycles

Throughout § 5.1 we fix  $\alpha \in \mathcal{O}_F$  and  $v \in F \otimes_{\mathbb{Q}} \mathbb{R}$ , both totally positive, and let  $v = (v_1, v_2)$  be the corresponding element under the isomorphism (3). Abbreviate  $\mathcal{Y} = \mathcal{Y}(\alpha)$  and  $\mathcal{Y}_0 = \mathcal{Y}_0(\alpha)$  as in earlier sections. For each prime  $p$  let

$$\mathcal{C}_p \in \text{CH}_{\mathcal{Y}/\mathbb{Z}_p}^2(\mathcal{M}/\mathbb{Z}_p)$$

be the cycle class (28) if  $p$  does not divide  $\text{disc}(B_0)$ , or the cycle class (38) if  $p$  does divide  $\text{disc}(B_0)$ . Let  $\mathcal{C}_{\mathbb{Q}} \in Z_{\mathcal{Y}/\mathbb{Q}}^2(\mathcal{M}/\mathbb{Q})$  be the cycle constructed in § 3.2 with Green current  $\Xi_v = \Xi_v(\alpha)$ . By Lemmas 3.3.2 and 4.2.5 each triple  $(\mathcal{C}_{\mathbb{Q}}, \Xi_v, \mathcal{C}_p)$  is a local cycle datum of codimension two with support on  $\mathcal{Y}$  in the sense of § 2.3, and it is easy to see that  $\mathcal{C}_p^{\text{ver}} = 0$  for any prime  $p$  with the property that  $\mathcal{Y}$  has no components supported in characteristic  $p$ . Hence  $(\mathcal{C}_{\mathbb{Q}}, \Xi_v, \mathcal{C}_{\bullet})$  is a global cycle datum. Applying the construction of (13) to  $(\mathcal{C}_{\mathbb{Q}}, \Xi_v, \mathcal{C}_{\bullet})$  yields an arithmetic cycle class

$$\widehat{\mathcal{Y}}(\alpha, v) = \widehat{\mathcal{Y}}(\alpha, v)^{\text{hor}} + \sum_p \widehat{\mathcal{Y}}(\alpha)_p^{\text{ver}} \in \widehat{\text{CH}}^2(\mathcal{M})$$

in which  $\widehat{\mathcal{Y}}(\alpha, v)^{\text{hor}}$  is the Zariski closure of  $\mathcal{C}_{\mathbb{Q}}$  endowed with the Green current  $\Xi_v$  and  $\widehat{\mathcal{Y}}(\alpha)_p^{\text{ver}}$  is the image of  $\mathcal{C}_p^{\text{ver}}$  under the map (11).

For the remainder of § 5.1 we assume that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic. By the final claim of Lemma 3.1.5 the stack  $\mathcal{Y}_0/\mathbb{Q}$  is empty, and so (15) allows us to compute the arithmetic degree along  $\mathcal{M}_0$  of the arithmetic cycle class  $\widehat{\mathcal{Y}}(\alpha, v)$  in terms of the global cycle datum  $(i^*\mathcal{C}_{\mathbb{Q}}, i^*\Xi_v, i^*\mathcal{C}_{\bullet})$  of codimension two supported on  $\mathcal{Y}_0$ . Indeed, (15) tells us that

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\mathcal{Y}}(\alpha, v) = \text{deg}_{\infty}(i^*\Xi_v) + \sum_{p \text{ prime}} \text{deg}_p(i^*\mathcal{C}_p) \cdot \log(p) \tag{45}$$

where, slightly abusing notation,  $\text{deg}_p$  is the composition

$$\text{CH}_{\mathcal{Y}_0/\mathbb{Z}_p}^2(\mathcal{M}_0/\mathbb{Z}_p) \rightarrow \text{CH}_{\text{ver}}^2(\mathcal{M}_0/\mathbb{Z}_p) \xrightarrow{\text{deg}_p} \mathbb{Q}.$$

For every  $T \in \text{Sym}_2(\mathbb{Z})^{\vee}$  and symmetric positive definite  $\mathbf{v} \in M_2(\mathbb{R})$ , Kudla, Rapoport and Yang have defined (see § 3.6 of [KRY06] and Chapter 6 of *loc. cit.*) an arithmetic zero cycle

$$\widehat{\mathcal{Z}}(T, \mathbf{v}) \in \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M}_0)$$

in the  $\mathbb{R}$ -arithmetic Chow group as defined in [KRY06, § 2.4]. Furthermore, when  $\det(T) \neq 0$  they define a finite set of places  $\text{Diff}(T, B_0)$  of  $\mathbb{Q}$  with the property

$$\mathcal{Z}(T)(\mathbb{F}_p^{\text{alg}}) \neq \emptyset \implies \text{Diff}(T, B_0) = \{p\} \tag{46}$$

for each finite place  $p$ . When  $T \in \text{Sym}_2(\mathbb{Z})^{\vee}$  is nonsingular and  $|\text{Diff}(T, B_0)| > 1$  then  $\widehat{\mathcal{Z}}(T, \mathbf{v}) = 0$  by definition. Recall that  $\det(T) \neq 0$  for every  $T \in \Sigma(\alpha)$  by Lemma 3.1.5 and also our hypothesis that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic.

Recalling our fixed  $\mathbb{Z}$ -basis  $\{\varpi_1, \varpi_2\}$  of  $\mathcal{O}_F$  set

$$R = \begin{pmatrix} \varpi_1 & \varpi_1^\sigma \\ \varpi_2 & \varpi_2^\sigma \end{pmatrix} \tag{47}$$

and abbreviate

$$\mathbf{v} = R \begin{pmatrix} v_1 & \\ & v_2 \end{pmatrix} {}^t R. \tag{48}$$

According to [KRY06, §2.4] the  $\mathbb{R}$ -arithmetic Chow group admits an *arithmetic degree* isomorphism

$$\widehat{\text{deg}} : \widehat{\text{CH}}_{\mathbb{R}}^2(\mathcal{M}) \cong \mathbb{R}$$

defined in exactly the same way as the arithmetic degree constructed in §2.3 for arithmetic Chow groups with rational coefficients.

LEMMA 5.1.1. *We have*

$$\text{deg}_\infty(i^* \Xi_v) = \sum_{\substack{T \in \Sigma(\alpha) \\ \text{Diff}(T, B_0) = \{\infty\}}} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}).$$

*Proof.* Return to the notation of §3.2. For any nonsingular  $T \in \text{Sym}_2(\mathbb{Z})^\vee$  we have  $\text{Diff}(T, B_0) = \{\infty\}$  if and only if the matrix  $T$  is represented by the quadratic space  $V_0$ . Assume that this is the case. For each pair  $(s_1, s_2) \in V_0(T)$  set

$$\tau = s_1 \varpi_1 + s_2 \varpi_2 \in V \cong V_0 \otimes_{\mathbb{Q}} F$$

and write  $(\tau_1, \tau_2)$  for the image of  $\tau$  under  $V \otimes_{\mathbb{Q}} \mathbb{R} \cong (V_0 \otimes_{\mathbb{Q}} \mathbb{R})^2$ . The matrix

$$\mathbf{u} = R \begin{pmatrix} v_1^{1/2} & \\ & v_2^{1/2} \end{pmatrix}$$

satisfies  $\mathbf{v} = \mathbf{u} \cdot {}^t \mathbf{u}$  and  $(s_1, s_2) \cdot \mathbf{u} = (v_1^{1/2} \tau_1, v_2^{1/2} \tau_2)$ . Therefore, the real number  $\nu_\infty(T, \mathbf{v})$  of [KRY06, Corollary 6.3.4] is given by

$$\nu_\infty(T, \mathbf{v}) = \frac{1}{2} \int_{X_0} \xi_0(v_1^{1/2} \tau_1) * \xi_0(v_2^{1/2} \tau_2)$$

and does not depend on the choice of  $(s_1, s_2) \in V_0(T)$ . But then by definition (see [KRY06, (6.3.2)]) we have

$$\widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) = \sum_{(s_1, s_2) \in \Gamma_0^{\max} \backslash L_0(T)} \frac{1}{2 \cdot |\text{stab}(s_1, s_2)|} \int_{X_0} \xi_0(v_1^{1/2} \tau_1) * \xi_0(v_2^{1/2} \tau_2) \tag{49}$$

where  $\text{stab}(s_1, s_2)$  is the stabilizer of  $(s_1, s_2)$  in  $\Gamma_0^{\max}$ . When  $\text{Diff}(T, B_0) \neq \{\infty\}$  the sum on the right-hand side of (49) is empty, and so the claim follows from Proposition 3.2.1.  $\square$

LEMMA 5.1.2. *If  $p \nmid \text{disc}(B_0)$  then*

$$\text{deg}_p(i^* \mathcal{C}_p) \log(p) = \sum_{\substack{T \in \Sigma(\alpha) \\ \text{Diff}(T, B_0) = \{p\}}} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}).$$



*Proof.* As  $p \nmid \text{disc}(B_0)$ , Proposition 6.11 and Corollary 6.1.2 of [KRY06] imply that, for every  $T \in \Sigma(\alpha)$  with  $\text{Diff}(T, B_0) = \{p\}$ ,

$$\widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) = \sum_{z \in \mathcal{Z}(T)(\mathbb{F}_p^{\text{alg}})} \frac{\log(p)}{|\text{Aut}_{\mathcal{Z}(T)}(z)|} \text{length}_{\mathcal{O}_{\mathcal{Z}(T),z}^{\text{sh}}}(\mathcal{O}_{\mathcal{Z}(T),z}^{\text{sh}})$$

where the sum is over the isomorphism classes of objects in  $\mathcal{Z}(T)(\mathbb{F}_p^{\text{alg}})$  and  $\mathcal{O}_{\mathcal{Z}(T),z}^{\text{sh}}$  is the strictly Henselian local ring of  $\mathcal{Z}(T)$  at  $z$ . On the other hand, as we assume that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic Lemma 3.1.5, the decomposition (21) and [KRY06, Theorem 3.6.1] imply that  $\mathcal{Y}_0/\mathbb{Z}_p$  is zero-dimensional. Thus we may apply Proposition 3.3.10 to obtain

$$\text{deg}_p(i^*C_p) = \sum_{y \in \mathcal{Y}_0(\mathbb{F}_p^{\text{alg}})} \frac{1}{|\text{Aut}_{\mathcal{Y}_0}(y)|} \text{length}_{\mathcal{O}_{\mathcal{Y}_0,y}^{\text{sh}}}(\mathcal{O}_{\mathcal{Y}_0,y}^{\text{sh}})$$

(by dividing both sides of the equality of Proposition 3.3.10 by  $|H|$ ). Combining this with the decomposition (21) and using (46) we arrive at the desired equality.  $\square$

LEMMA 5.1.3. *Suppose  $p \mid \text{disc}(B_0)$ . Then*

$$\text{deg}_p(i^*C_p) \log(p) = \sum_{\substack{T \in \Sigma(\alpha) \\ \text{Diff}(T, B_0) = \{p\}}} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}).$$

*Proof.* Return to the notation of § 4. By Corollary 4.3.3 we have

$$\text{deg}_p(i^*C_p) = \chi(\mathcal{O}_{\mathfrak{m}_0^1} \otimes_{\mathcal{O}_{\mathfrak{m}_0}}^L \mathcal{O}_{\mathfrak{m}_0^2}).$$

For each  $T \in \Sigma(\alpha)$  either  $\text{Diff}(T, B_0) \neq \{p\}$  or  $\text{Diff}(T, B_0) = \{p\}$ . In the former case  $\mathcal{Z}(T)(\mathbb{F}_p^{\text{alg}}) = \emptyset$ . In the latter case the matrix  $T$  is represented by the quadratic space  $\overline{V}_0$  of § 4.1, and so the hypotheses of Proposition 4.3.4 are satisfied. Therefore,

$$\frac{1}{|H|} \cdot \text{deg}_p(i^*C_p) = \sum_{\substack{T \in \Sigma(\alpha) \\ \text{Diff}(T, B_0) = \{p\}}} e_p(T) \cdot |\overline{\Gamma}_0 \backslash \overline{L}_0(T)|.$$

By [KRY06, Theorem 6.2.1] we have, for any  $T$  appearing in the sum,

$$\widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) = e_p(T) \cdot |\overline{\Gamma}_0 \backslash \overline{L}_0(T)| \cdot \log(p)$$

which completes the proof.  $\square$

THEOREM 5.1.4. *Suppose we are given  $\alpha \in \mathcal{O}_F$  and  $v \in F \otimes_{\mathbb{Q}} \mathbb{R}$ , both totally positive. If the extension  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic then*

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\mathcal{Y}}(\alpha, v) = \sum_{T \in \Sigma(\alpha)} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v})$$

where  $\mathbf{v}$  is defined by (48).

*Proof.* This is immediate from the decomposition (45), Lemmas 5.1.1, 5.1.2, and 5.1.3, and, by definition of  $\widehat{\mathcal{Z}}(T, \mathbf{v})$ , the implication

$$|\text{Diff}(T, B_0)| > 1 \implies \widehat{\mathcal{Z}}(T, \mathbf{v}) = 0$$

for each nonsingular  $T \in \text{Sym}_2(\mathbb{Z})^\vee$ .  $\square$

5.2 A genus two Eisenstein series

Let  $J_n$  be the  $2n \times 2n$  matrix

$$J_n = \begin{pmatrix} & I_n \\ -I_n & \end{pmatrix}.$$

Recall our fixed basis  $\{\varpi_1, \varpi_2\}$  for the  $\mathbb{Z}$ -module  $\mathcal{O}_F$  and let  $\{\varpi_1^\vee, \varpi_2^\vee\}$  be the dual basis relative to the trace form  $(x, y) \mapsto \text{Tr}_{F/\mathbb{Q}}(xy)$ . The  $F$ -vector space  $V_F = F^2$  comes equipped with the standard  $F$ -valued symplectic form  $\langle v, w \rangle_F = {}^t w J_1 v$ , while the  $\mathbb{Q}$ -vector space  $V_{\mathbb{Q}}$  underlying  $V_F$  is equipped with the symplectic form

$$\langle v, w \rangle_{\mathbb{Q}} = \text{Tr}_{F/\mathbb{Q}} \langle v, w \rangle_F.$$

With respect to the  $\mathbb{Q}$ -basis

$$\left\{ \begin{bmatrix} \varpi_1^\vee \\ 0 \end{bmatrix}, \begin{bmatrix} \varpi_2^\vee \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \varpi_1 \end{bmatrix}, \begin{bmatrix} 0 \\ \varpi_2 \end{bmatrix} \right\}$$

of  $V_{\mathbb{Q}}$ , the symplectic form  $\langle v, w \rangle_{\mathbb{Q}}$  is given by  $\langle v, w \rangle_{\mathbb{Q}} = {}^t w J_2 v$ . In this way we obtain a homomorphism

$$\text{Sp}_1(F) \cong \text{Sp}(V_F) \hookrightarrow \text{Sp}(V_{\mathbb{Q}}) \cong \text{Sp}_2(\mathbb{Q})$$

which is given explicitly by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} R & \\ & {}^t R^{-1} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} R^{-1} & \\ & {}^t R \end{pmatrix}$$

where  $R$  is defined by (47) and

$$\underline{x} = \begin{pmatrix} x \\ x^\sigma \end{pmatrix}$$

for  $x \in F$ . If  $\Gamma_0(M) \subset \text{Sp}_2(\mathbb{Z})$  is the usual subgroup of matrices congruent modulo  $M$  to  $\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$  then

$$\text{Sp}_1(F) \cap \Gamma_0(M) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Sp}_1(F) \mid a, d \in \mathcal{O}_F, b \in \mathfrak{D}^{-1}, c \in M\mathfrak{D} \right\}.$$

The group  $\text{Sp}_1(F)$  acts in the usual way on the product of two upper half-planes  $\mathfrak{h}_1 \times \mathfrak{h}_1$  via the embedding

$$\text{Sp}_1(F) \hookrightarrow \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$$

defined by  $A \mapsto (A, A^\sigma)$ . Similarly, the rank two symplectic group  $\text{Sp}_2(\mathbb{Q})$  acts in the habitual manner on the Siegel half-space  $\mathfrak{h}_2$  of genus two. If we let  $\text{Sp}_1(F)$  act on  $\mathfrak{h}_2$  through the above homomorphism  $\text{Sp}_1(F) \rightarrow \text{Sp}_2(\mathbb{Q})$  then the embedding of complex manifolds

$$\mathfrak{h}_1 \times \mathfrak{h}_1 \hookrightarrow \mathfrak{h}_2, \quad (\tau_1, \tau_2) \mapsto R \begin{pmatrix} \tau_1 & \\ & \tau_2 \end{pmatrix} {}^t R \tag{50}$$

respects the actions of  $\text{Sp}_1(F)$ .

To the quaternionic order  $\mathcal{O}_{B_0}$ , Kudla [Kud97] attaches a Siegel Eisenstein series of weight  $3/2$ ,

$$\mathcal{E}_2(\tau, s, B_0) = \sum_{T \in \text{Sym}_2(\mathbb{Q})} \mathcal{E}_{2,T}(\tau, s, B_0),$$

where  $\tau = \mathbf{u} + i\mathbf{v} \in \mathfrak{h}_2$  and  $s \in \mathbb{C}$ . See [KRY06, (5.1.44)] for the precise definition. The Eisenstein series satisfies a function equation which forces  $\mathcal{E}_{2,T}(\tau, 0, B_0) = 0$  for every  $T$ , and one of the

main results of Kudla, Rapoport and Yang [KRY06, p. 12, Theorem B] is the equality

$$\frac{d}{ds} \mathcal{E}_{2,T}(\tau, s, B_0)|_{s=0} = \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \cdot q^T$$

for every  $T \in \text{Sym}_2(\mathbb{Q})$ , in which  $q^T = e^{2\pi i \cdot \text{Tr}(T\tau)}$  and

$$\widehat{\mathcal{Z}}(T, \mathbf{v}) \in \widehat{\text{CH}}^2(\mathcal{M}_0)$$

is the arithmetic cycle class appearing in § 5.1 and defined in [KRY06, ch. 6]. When  $T \notin \text{Sym}_2(\mathbb{Z})^\vee$ , both sides of the stated equality are zero. In particular, if we denote by  $\widehat{\phi}_2(\tau)$  Kudla's  $q$ -expansion

$$\widehat{\phi}_2(\tau) = \sum_{T \in \text{Sym}_2(\mathbb{Z})^\vee} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \cdot q^T$$

then  $\mathcal{E}'_2(\tau, 0, B_0) = \widehat{\phi}_2(\tau)$  is a (non-holomorphic) Siegel modular form of weight  $3/2$  for the congruence subgroup  $\Gamma_0(4m_o) \subset \text{Sp}_2(\mathbb{Z})$  where

$$m_o = \text{disc}(B_0) \cdot \begin{cases} 1 & \text{if } \text{disc}(B_0) \text{ is odd,} \\ 1/2 & \text{if } \text{disc}(B_0) \text{ is even.} \end{cases}$$

It follows that the pullback of  $\widehat{\phi}_2$  to  $\mathfrak{h}_1 \times \mathfrak{h}_1$  is a Hilbert modular form of parallel weight  $3/2$  for the congruence subgroup

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(F) \mid a, d \in \mathcal{O}_F, b \in \mathfrak{D}^{-1}, c \in 4m_o\mathfrak{D} \right\}. \tag{51}$$

LEMMA 5.2.1. *The pullback of  $\widehat{\phi}_2$  to  $\mathfrak{h}_1 \times \mathfrak{h}_1$  has Fourier expansion*

$$\widehat{\phi}_2(\tau_1, \tau_2) = \sum_{\alpha \in \mathcal{O}_F} \left( \sum_{T \in \Sigma(\alpha)} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \right) \cdot q^\alpha$$

where we have set  $q^\alpha = e^{2\pi i \tau_1 \alpha} \cdot e^{2\pi i \tau_2 \alpha^\sigma}$ . The positive definite matrix  $\mathbf{v}$  defined by (48), with  $v_1$  and  $v_2$  the imaginary parts of  $\tau_1$  and  $\tau_2$  respectively, is the imaginary part of the image of  $(\tau_1, \tau_2)$  in  $\mathfrak{h}_2$ .

*Proof.* If we set  $\tau = R \begin{pmatrix} \tau_1 & \\ & \tau_2 \end{pmatrix} {}^t R$  then for every  $\alpha \in \mathcal{O}_F$  and  $T \in \Sigma(\alpha)$  we have

$$\text{Tr}(T\tau) = \alpha\tau_1 + \alpha^\sigma\tau_2.$$

It follows now from  $\text{Sym}_2(\mathbb{Z})^\vee = \bigcup_{\alpha \in \mathcal{O}_F} \Sigma(\alpha)$  that

$$\widehat{\phi}_2(\tau_1, \tau_2) = \sum_{\alpha \in \mathcal{O}_F} \sum_{T \in \Sigma(\alpha)} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v}) \cdot e^{2\pi i \alpha \tau_1} e^{2\pi i \alpha^\sigma \tau_2},$$

proving the claim. □

COROLLARY 5.2.2. *Suppose  $\alpha \in \mathcal{O}_F$  is totally positive and that  $F(\sqrt{-\alpha})/\mathbb{Q}$  is not biquadratic. Then the  $\alpha$ th Fourier coefficient of  $\widehat{\phi}_2(\tau_1, \tau_2)$  is*

$$\widehat{\text{deg}}_{\mathcal{M}_0} \widehat{\mathcal{Y}}(\alpha, v) = \sum_{T \in \Sigma(\alpha)} \widehat{\text{deg}} \widehat{\mathcal{Z}}(T, \mathbf{v})$$

where  $v = (v_1, v_2) \in \mathbb{R} \times \mathbb{R}$  is the imaginary part of  $(\tau_1, \tau_2)$  and  $\mathbf{v}$  is defined by (48).

*Proof.* This is immediate from Lemma 5.2.1 and Theorem 5.1.4. □

## ACKNOWLEDGEMENT

The author thanks Steve Kudla for suggesting this problem, and for several helpful conversations.

## REFERENCES

- AG03 F. Andreatta and E. Z. Goren, *Geometry of Hilbert modular varieties over totally ramified primes*, Int. Math. Res. Not. **33** (2003), 1786–1835.
- AG04 F. Andreatta and E. Z. Goren, *Hilbert modular varieties of low dimension*, Geometric Aspects of Dwork Theory. vol. I, II (Walter de Gruyter GmbH & Co. KG, Berlin, 2004), 113–175.
- AG05 F. Andreatta and E. Z. Goren, *Hilbert modular forms: mod  $p$  and  $p$ -adic aspects*, Mem. Amer. Math. Soc. **173** (2005), vi+100.
- BLR90 S. Bosch, W. Lütkebohmert and M. Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 21 (Springer-Verlag, Berlin, 1990).
- Bou97 J.-F. Boutot, *Uniformisation  $p$ -adique des variétés de Shimura*, Astérisque **245** (1997), 307–322.
- BBGGLMS79 J.-F. Boutot, L. Breen, P. Gérardin, J. Giraud, J.-P. Labesse, J. Milne and C. Soulé, *Variétés de Shimura et fonctions  $L$* , Publications Mathématiques de l'Université Paris VII [Mathematical Publications of the University of Paris VII], vol. 6 (Université de Paris VII U.E.R. de Mathématiques, Paris, 1979).
- BC92 J.-F. Boutot and H. Carayol, *Uniformisation  $p$ -adique des courbes de Shimura: les théorèmes de Čerednik et de Drinfel'd*, Astérisque **196–197** (1991), 45–158. (Courbes modulaires et courbes de Shimura (Orsay, 1987/1988).)
- BZ J.-F. Boutot and T. Zink, *The  $p$ -adic uniformization of Shimura curves*. Preprint, <http://www.mathematik.uni-bielefeld.de/~zink>.
- BBK07 Jan H. Bruinier, J. I. Burgos Gil and U. Kühn, *Borchers products and arithmetic intersection theory on Hilbert modular surfaces*, Duke Math. J. **139** (2007), 1–88.
- BKK07 J. I. Burgos Gil, J. Kramer and U. Kühn, *Cohomological arithmetic Chow rings*, J. Inst. Math. Jussieu **6** (2007), 1–172.
- Buz97 K. Buzzard, *Integral models of certain Shimura curves*, Duke Math. J. **87** (1997), 591–612.
- Con04 B. Conrad, *Gross-Zagier revisited*, in *Heegner Points and Rankin  $L$ -series*, Mathematical Science Research Institute Publications, vol. 49 (Cambridge University Press, Cambridge, 2004), 67–163. (With an appendix by W. R. Mann.)
- DP94 P. Deligne and G. Pappas, *Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant*, Compositio Math. **90** (1994), 59–79.
- FGIKNV05 B. Fantechi, L. Göttsche, L. Illusie, S. L. Kleiman, N. Nitsure and A. Vistoli, *Fundamental algebraic geometry: Grothendieck's FGA explained*, Mathematical Surveys and Monographs, vol. 123 (American Mathematical Society, Providence, RI, 2005).
- Ful98 W. Fulton, *Intersection theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 39, second edition (Springer, Berlin, 1998).
- Gil84 H. Gillet, *Intersection theory on algebraic stacks and  $Q$ -varieties*, J. Pure Appl. Algebra **34** (1984), 193–240.
- GS87 H. Gillet and C. Soulé, *Intersection theory using Adams operations*, Invent. Math. **90** (1987), 243–277.
- GS91 H. Gillet and C. Soulé, *Arithmetic intersection theory*, Publ. Math. Inst. Hautes Études Sci. **72** (1990), 93–174.

- Gor02 E. Z. Goren, *Lectures on Hilbert modular varieties and modular forms*, CRM Monograph Series, vol. 14 (American Mathematical Society, Providence, RI, 2002). (With the assistance of Marc-Hubert Nicole.)
- GO00 E. Z. Goren and F. Oort, *Stratifications of Hilbert modular varieties*, *J. Algebraic Geom.* **9** (2000), 111–154.
- Gub02 W. Gubler, *Moving lemma for  $K_1$ -chains*, *J. Reine Angew. Math.* **548** (2002), 1–19.
- Hid04 H. Hida,  *$p$ -adic automorphic forms on Shimura varieties*, Springer Monographs in Mathematics (Springer-Verlag, New York, 2004).
- Hid06 H. Hida, *Hilbert modular forms and Iwasawa theory*, Oxford Mathematical Monographs (The Clarendon Press; Oxford University Press, Oxford, 2006).
- How B. Howard, *Intersection theory on Shimura surfaces II*. Preprint.
- Kot92 R. Kottwitz, *Points on some Shimura varieties over finite fields*, *J. Amer. Math. Soc.* **5** (1992), 373–444.
- Kud97 S. Kudla, *Central derivatives of Eisenstein series and height pairings*, *Ann. of Math. (2)* **146** (1997), 545–646.
- Kud04 S. Kudla, *Special cycles and derivatives of Eisenstein series*, in *Heegner points and Rankin  $L$ -series*, Mathematical Science Research Institute Publications, vol. 49 (Cambridge University Press, Cambridge, 2004), 243–270.
- KR99 S. Kudla and M. Rapoport, *Arithmetic Hirzebruch-Zagier cycles*, *J. Reine Angew. Math.* **515** (1999), 155–244.
- KR00 S. Kudla and M. Rapoport, *Height pairings on Shimura curves and  $p$ -adic uniformization*, *Invent. Math.* **142** (2000), 153–223.
- KRY04 S. Kudla, M. Rapoport and T. Yang, *Derivatives of Eisenstein series and Faltings heights*, *Composito Math.* **140** (2004), 887–951.
- KRY06 S. Kudla, M. Rapoport and T. Yang, *Modular forms and special cycles on Shimura curves*, *Annals of Mathematics Studies*, vol. 161 (Princeton University Press, Princeton, NJ, 2006).
- Lam05 T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics, vol. 67 (American Mathematical Society, Providence, RI, 2005).
- LM00 G. Laumon and L. Moret-Bailly, *Champs algébriques*, *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*, vol. 39 (Springer, Berlin, 2000).
- Liu02 Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics, vol. 6 (Oxford Science Publications, Oxford University Press, Oxford, 2002). (Translated from the French by Reinie Ern .)
- Mat89 H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics, vol. 8 second edition (Cambridge University Press, Cambridge, 1989). (Translated from the Japanese by M. Reid.)
- Mil90 J. Milne, *Canonical models of (mixed) Shimura varieties and automorphic vector bundles*, in *Automorphic forms, Shimura varieties, and  $L$ -functions, vol. I*, (Ann Arbor, MI, 1988, Perspectives in Mathematics, vol. 10 (Academic Press, Boston, MA, 1990), 283–414.
- Mil05 J. Milne, *Introduction to Shimura varieties*, in *Harmonic analysis, the trace formula, and Shimura varieties*, Clay Mathematical Proceedings, vol. 4 (American Mathematical Society, Providence, RI, 2005), 265–378.
- Mil79 J. S. Milne, *Points on Shimura varieties mod  $p$* , in *Automorphic forms, representations and  $L$ -functions*, Oregon State University, Corvallis, OR, 1977, Part 2, Proceedings of Symposia in Pure Mathematics, XXXIII (American Mathematical Society, Providence, RI, 1979), 165–184.

- Mum65 D. Mumford, *Geometric invariant theory*, in *Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34* (Springer, Berlin, 1965).
- Pap95 G. Pappas, *Arithmetic models for Hilbert modular varieties*, *Compositio Math.* **98** (1995), 43–76.
- Rap78 M. Rapoport, *Compactifications de l'espace de modules de Hilbert–Blumenthal*, *Compositio Math.* **36** (1978), 255–335.
- RZ82 M. Rapoport and T. Zink, *Über die lokale Zetafunktion von Shimuravarietäten. Monodromiefiltration und verschwindende Zyklen in ungleicher Charakteristik*, *Invent. Math.* **68** (1982), 21–101.
- RZ96 M. Rapoport and Th. Zink, *Period spaces for  $p$ -divisible groups*, *Annals of Mathematics Studies*, vol. 141 (Princeton University Press, Princeton, NJ, 1996).
- Ser00 J.-P. Serre, *Local algebra*, Springer Monographs in Mathematics (Springer, Berlin, 2000). (Translated from the French by CheeWhye Chin and revised by the author.)
- Sou92 C. Soulé, *Lectures on Arakelov geometry*, Cambridge Studies in Advanced Mathematics, vol. 33 (Cambridge University Press, Cambridge, 1992). (With the collaboration of D. Abramovich, J.-F. Burnol and J. Kramer.)
- Sta97 H. Stamm, *On the reduction of the Hilbert–Blumenthal-moduli scheme with  $\Gamma_0(p)$ -level structure*, *Forum Math.* **9** (1997), 405–455.
- van88 G. van der Geer, *Hilbert modular surfaces*, *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, vol. 16 (Springer, Berlin, 1988).
- Vis89 A. Vistoli, *Intersection theory on algebraic stacks and on their moduli spaces*, *Invent. Math.* **97** (1989), 613–670.
- Vol05 I. Vollaard, *On the Hilbert–Blumenthal moduli problem*, *J. Inst. Math. Jussieu* **4** (2005), 653–683.
- Wew07 S. Wewers, *Canonical and quasi-canonical liftings*, *Astérisque* **312** (2007), 67–86.
- Zin02 T. Zink, *The display of a formal  $p$ -divisible group*, *Astérisque* **278** (2002), 127–248.

Benjamin Howard [howardbe@bc.edu](mailto:howardbe@bc.edu)

Department of Mathematics, Boston College, Chestnut Hill, MA, 02467, USA