Algebra & Number Theory

Volume 11
2017
No. 10

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2017 is US $325/year for the electronic version, and $520/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2213

Tate cycles on some unitary Shimura varieties mod p

David Helm, Yichao Tian and Liang Xiao

Let F be a real quadratic field in which a fixed prime p is inert, and E_0 be an imaginary quadratic field in which p splits; put $E = E_0F$. Let X be the fiber over $\mathbb{F}_{p^2}$ of the Shimura variety for $G(U(1, n-1) \times U(n-1, 1))$ with hyperspecial level structure at p for some integer $n \geq 2$. We show that under some genericity conditions the middle-dimensional Tate classes of X are generated by the irreducible components of its supersingular locus. We also discuss a general conjecture regarding special cycles on the special fibers of unitary Shimura varieties, and on their relation to Newton stratification.

1. Introduction

The study of the geometry of Shimura varieties lies at the heart of the Langlands program. Arithmetic information of Shimura varieties builds a bridge relating the world of automorphic representations and the world of Galois representations.

One of the interesting topics in this area is to understand the supersingular locus of the special fibers of Shimura varieties, or more generally, any interesting stratifications (e.g., Newton or Ekedahl–Oort stratification) of the special fibers of Shimura varieties. The case of unitary Shimura varieties has been extensively

MSC2010: primary 11G18; secondary 11R39, 14C17, 14C25, 14G35.

Keywords: Supersingular locus, Special fiber of Shimura varieties, Deligne–Lusztig varieties, Tate conjecture.

studied. Vollaard and Wedhorn [2011] showed that the supersingular locus of the special fiber of the $GU(1, n-1)$-Shimura variety at an inert prime is a union of Deligne–Lusztig varieties. Further, Howard and Pappas [2014] studied the case of $GU(2, 2)$ at an inert prime, and Rapoport, Terstiege and Wilson proved similar results for $GU(n-1, 1)$ at a ramified prime. Finally, we remark that Görtz and He [2015] studied the basic loci in a slightly more general class of Shimura varieties.

In all the work mentioned above, the authors use the uniformization theorem of Rapoport–Zink to reduce the problem to the study of certain Rapoport–Zink spaces. In this paper, we take a different approach. Instead of using the uniformization theorem, we study the basic locus (or more generally other Newton strata) of certain unitary Shimura varieties by considering correspondences between unitary Shimura varieties of different signatures. This method was introduced by the first author in [Helm 2010; 2012], and applied successfully to quaternionic Shimura varieties by the second and the third authors [Tian and Xiao 2016].

Another new aspect of this work is that we study not only the global geometry of the supersingular locus, but also their relationship with the Tate conjecture for Shimura varieties over finite fields. We show that the basic locus contributes to all "generic" middle-dimensional Tate cycles of the special fiber of the Shimura variety. Similar results have been obtained by the second and the third authors for even-dimensional Hilbert modular varieties at an inert prime [Tian and Xiao 2014]. We believe that, this phenomenon is a general philosophy which holds for more general Shimura varieties. Our slogan is: *irreducible components of the basic locus of a Shimura variety should generate all Tate classes under some genericity condition on the automorphic representations.*

We explain in more detail the main results of this paper. Let F be a real quadratic field, E_0 be an imaginary quadratic field, and $E = E_0F$. Let p be a prime number inert in F, and split in E_0. Let $\mathfrak{p}$, $\bar{\mathfrak{p}}$ denote the two places of E above p so that $E_{\mathfrak{p}}$ and $E_{\bar{\mathfrak{p}}}$ are both isomorphic to $\mathbb{Q}_{p^2}$, the unique unramified quadratic extension of $\mathbb{Q}_p$. For an integer $n \geq 1$, let G be the similitude unitary group associated to a division algebra over E equipped with an involution of second kind. In the notation of Section 3.6, our G is denoted $G_{1,n-1}$. This is an algebraic group over $\mathbb{Q}$ such that $G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_{\mathfrak{p}})$ and $G(\mathbb{R})$ is the unitary similitude group with signature $(1, n-1)$ and $(n-1, 1)$ at the two archimedean places. (For a precise definition, see Section 2.2.)

Let $\mathbb{A}$ denote the ring of finite adeles of $\mathbb{Q}$, and $\mathbb{A}^\infty$ be its finite part. Fix a sufficiently small open compact subgroup $K \subseteq G(\mathbb{A}^\infty)$ with $K_p = \mathbb{Z}_p^\times \times \mathrm{GL}_n(\mathbb{Z}_{p^2}) \subseteq G(\mathbb{Q}_p)$, where $\mathbb{Z}_{p^2}$ is the ring of integers of $\mathbb{Q}_{p^2}$. Let $\mathcal{S}h(G)_K$ be the Shimura variety associated to G of level K.[1]

[1] Strictly speaking, the moduli space $\mathcal{S}h(G)_K$ is $\#\ker^1(\mathbb{Q}, G)$-copies of the classical Shimura variety whose $\mathbb{C}$-points are given by the double coset space $G(\mathbb{Q}) \setminus G(\mathbb{A})/K_\infty K$, where $K_\infty \subseteq G(\mathbb{R})$ is the maximal compact subgroup modulo center. See [Kottwitz 1992b, page 400] for details.

According to Kottwitz [1992b], when K^p is neat, $\mathcal{S}h(G)_K$ admits a proper and smooth integral model over $\mathbb{Z}_{p^2}$ which parametrizes certain polarized abelian schemes with K-level structure (See Section 2.3). Let $\mathrm{Sh}_{1,n-1}$ denote the special fiber of $\mathcal{S}h(G)_K$ over $\mathbb{F}_{p^2}$. This is a proper smooth variety over $\mathbb{F}_{p^2}$ of dimension $2(n-1)$. Let $\mathrm{Sh}^{\mathrm{ss}}_{1,n-1}$ denote the supersingular locus of $\mathrm{Sh}_{1,n-1}$, i.e., the reduced closed subvariety of $\mathrm{Sh}_{1,n-1}$ that parametrizes supersingular abelian varieties. We will see in Proposition 4.14 that $\mathrm{Sh}^{\mathrm{ss}}_{1,n-1}$ is equidimensional of dimension $n-1$.

Fix a prime $\ell \neq p$. There is a natural action by $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^2}) \times \overline{\mathbb{Q}}_\ell[K\backslash G(\mathbb{A}^\infty)/K]$ on the ℓ-adic étale cohomology group $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))$. We will take advantage of the Hecke action to consider a variant of the Tate conjecture for $\mathrm{Sh}_{1,n-1}$.

Fix an irreducible admissible representation π of $G(\mathbb{A}^\infty)$ (with coefficients in $\overline{\mathbb{Q}}_\ell$). The K-invariant subspace of π, denoted by π^K, is a finite-dimensional irreducible representation of the Hecke algebra $\overline{\mathbb{Q}}_\ell[K\backslash G(\mathbb{A}^\infty)/K]$. We denote the π^K-isotypic component of $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))$ by $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))_\pi$ and put

$$H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi := \bigcup_{\mathbb{F}_q/\mathbb{F}_{p^2}} H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_q)}_\pi,$$

where $\mathbb{F}_q$ runs through all finite extensions of $\mathbb{F}_{p^2}$. By projecting to the π^K-isotypic component, we have an ℓ-adic cycle class map:

$$\mathrm{cl}^{n-1}_\pi : A^{n-1}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}) \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}_\ell \to H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi, \tag{1.0.1}$$

where $A^{n-1}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p})$ is the abelian group of codimension $n-1$ algebraic cycles on $\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}$. Then the Tate conjecture for $\mathrm{Sh}_{1,n-1}$ predicts that the above map is surjective. Our main result confirms exactly this statement under some "genericity" assumptions on π.

From now on, we assume that π satisfies Hypothesis 2.5 to ensure the nontriviality of the π-isotypic component of the cohomology groups. In particular, π is the finite part of an automorphic cuspidal representation of $G(\mathbb{A})$, and $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))_\pi \neq 0$. Let π_p denote the p-component of π, which is an unramified principal series as K_p is hyperspecial. Since $G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_\mathfrak{p})$, we write $\pi_p = \pi_{p,0} \otimes \pi_\mathfrak{p}$, where $\pi_{p,0}$ is a character of $\mathbb{Q}_p^\times$ and $\pi_\mathfrak{p}$ is an irreducible admissible representation of $\mathrm{GL}_n(E_\mathfrak{p})$.

Our main theorem is the following.

Theorem 1.1. *Suppose π is the finite part of an automorphic representation of $G(\mathbb{A})$ that admits a cuspidal base change to $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}^\times_{E_0}$, and the Satake parameters of $\pi_\mathfrak{p}$ are distinct modulo roots of unity. Then $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi$ is generated by the cohomological classes of the irreducible components of the supersingular locus $\mathrm{Sh}^{\mathrm{ss}}_{1,n-1}$. In particular, the cycle class map* (1.0.1) *is surjective.*

This theorem will be restated in a more precise form in Theorem 4.18. Here, the assumption that the Satake parameters of $\pi_{\mathfrak{p}}$ are distinct modulo roots of unity is crucial for our method. It is closely tied to our geometric description of the irreducible components. This condition will be reformulated in Theorem 4.18 in terms of the Frobenius eigenvalues of certain Galois representation attached to $\pi_{\mathfrak{p}}$ via the unramified local Langlands correspondence. The other automorphic assumption on π is of technical nature. It is imposed here to ensure certain equalities on the automorphic multiplicity on π (See Remark 4.19). The method of our paper may be extended to more general representations π if we have more knowledge of the multiplicity of automorphic forms on unitary groups.

What we will prove is more precise than stated in Theorem 1.1. We need another unitary group $G' = G_{0,n}$ over $\mathbb{Q}$ for E/F as in Lemma 2.9, which is the unique inner form of G such that $G'(\mathbb{A}^\infty) \simeq G(\mathbb{A}^\infty)$ and the signatures of G' at the two archimedean places are $(0, n)$ and $(n, 0)$. Let $\mathrm{Sh}_{0,n}$ denote the (zero-dimensional) Shimura variety over $\mathbb{F}_{p^2}$ associated to G'. We will show in Proposition 4.14 that the supersingular locus $\mathrm{Sh}_{1,n-1}^{\mathrm{ss}}$ is a union of n closed subvarieties Y_j with $1 \le j \le n$ such that each of Y_j admits a fibration over $\mathrm{Sh}_{0,n}$ of the same level $K \subseteq G(\mathbb{A}^\infty) \simeq G'(\mathbb{A}^\infty)$ with fibers isomorphic to a certain proper and smooth closed subvariety in a product of Grassmannians. In other words, each Y_j is an algebraic correspondence between $\mathrm{Sh}_{1,n-1}$ and $\mathrm{Sh}_{0,n}$:

$$\mathrm{Sh}_{0,n} \leftarrow Y_j \rightarrow \mathrm{Sh}_{1,n-1}\,.$$

This can be viewed as a geometric realization of the Jacquet–Langlands correspondence between G and G' in the sense of [Helm 2010]. Alternatively, we may view these Y_j as Hecke correspondences between special fibers of unitary Shimura varieties of different signatures. To prove Theorem 1.1, it suffices to show that, when the Satake parameters of $\pi_{\mathfrak{p}}$ are distinct modulo roots of unity, $H^{2(n-1)}_{\mathrm{et}}(\mathrm{Sh}_{1,n-1,\bar{\mathbb{F}}_p}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi$ is generated by the cohomology classes of the irreducible components of Y_j. The key point is to show that the π-projection of the intersection matrix of Y_j is nondegenerate under the assumption above on $\pi_{\mathfrak{p}}$.

We briefly describe the structure of this paper. In Section 2, we consider a more general setup of unitary Shimura varieties, and propose a general conjecture, which roughly predicts the existence of certain algebraic correspondences between the special fibers of Shimura varieties with hyperspecial level at p associated to unitary groups with different signatures at infinity (Conjecture 2.12). Theorem 1.1 is a special case of Conjecture 2.12. We believe that our conjecture will provide a new perspective to understand the special fibers of Shimura varieties. In Section 3, we review some Dieudonné theory and Grothendieck–Messing deformation theory that will be frequently used in later sections. Section 4 is devoted to the study of the supersingular locus $\mathrm{Sh}_{1,n-1}^{\mathrm{ss}}$, and constructing the subvarieties Y_j mentioned above. In Section 5, we compute certain intersection numbers on products of Grassmannian varieties. These numbers will play a fundamental role in our later computation of the intersection matrix of the Y_j. In Section 6, we will compute explicitly the intersection matrix of the Y_j (Theorem 6.7), and show that its π-isotypic projection of the intersection matrix is nondegenerate as long as the Satake parameters of $\pi_\mathfrak{p}$ are distinct (as opposed to being distinct modulo roots of unity). Then an easy cohomological computation allows us to conclude the proof of our main theorem. In Section 7, we will generalize the construction of the cycles Y_j to the Shimura variety associated to unitary group for E/F of signature $(r, s) \times (s, r)$ at infinity. In this case, we only obtain some partial results on these cycles predicted by Conjecture 2.12: the union of these cycles is exactly the supersingular locus of the unitary Shimura variety in question (Theorem 7.8).

2. The conjecture on special cycles

We will only discuss certain unitary Shimura varieties so that the description becomes explicit. We will discuss after Conjecture 2.12 on how to possibly extend this conjecture to more general Shimura varieties.

2.1. ***Notation.*** We fix a prime number p throughout this paper. We fix an isomorphism $\iota_p : \mathbb{C} \xrightarrow{\sim} \overline{\mathbb{Q}}_p$. Let $\mathbb{Q}_p^{\mathrm{ur}}$ be the maximal unramified extension of $\mathbb{Q}_p$ inside $\overline{\mathbb{Q}}_p$.

Let F be a totally real field of degree f in which p is inert. We label all real embeddings of F, or equivalently (via ι_p), all p-adic embeddings of F (into $\mathbb{Q}_p^{\mathrm{ur}}$) by $\tau_1, \dots, \tau_f$ so that post-composition by the Frobenius map takes τ_i to τ_{i+1}. Here the subindices are taken modulo f. Let E_0 be an imaginary quadratic extension of $\mathbb{Q}$ in which p splits. Put $E = E_0 F$. Denote by v and $\bar{v}$ the two p-adic places of E_0. Then p splits into two primes $\mathfrak{p}$ and $\bar{\mathfrak{p}}$ in E, where $\mathfrak{p}$ (resp. $\bar{\mathfrak{p}}$) is the p-adic place above v (resp. $\bar{v}$). Let q_i denote the embedding $E \to E_\mathfrak{p} \cong F_p \xrightarrow{\tau_i} \overline{\mathbb{Q}}_p$ and $\bar{q}_i$ the analogous embedding which factors through $E_{\bar{\mathfrak{p}}}$ instead. Composing with ι_p^{-1}, we regard q_i and $\bar{q}_i$ as complex embeddings of E, and we put $\Sigma_{\infty,E} = \{q_1, \dots, q_f, \bar{q}_1, \dots, \bar{q}_f\}$.

2.2. ***Shimura data.*** Let D be a division algebra of dimension n^2 over its center E, equipped with a positive involution $*$ which restricts to the complex conjugation c on E. In particular, $D^{\mathrm{opp}} \cong D \otimes_{E,c} E$. We assume that D splits at $\mathfrak{p}$ and $\bar{\mathfrak{p}}$, and we fix an isomorphism

$$D \otimes_\mathbb{Q} \mathbb{Q}_p \simeq \mathrm{M}_n(E_\mathfrak{p}) \times \mathrm{M}_n(E_{\bar{\mathfrak{p}}}) \cong \mathrm{M}_n(\mathbb{Q}_{p^f}) \times \mathrm{M}_n(\mathbb{Q}_{p^f}),$$

where $*$ switches the two direct factors. We use $\mathfrak{e}$ to denote the element of $D \otimes_\mathbb{Q} \mathbb{Q}_p$ corresponding to the $(1, 1)$-elementary matrix[2] in the first factor. Let $a_\bullet = (a_i)_{1 \le i \le f}$

[2]By a $(1, 1)$-elementary matrix, we mean an $n \times n$-matrix whose $(1, 1)$-entry is 1 and whose other entries are zero.

be a tuple of f numbers with $a_i \in \{0, \dots, n\}$. Assume that there exists an element $\beta_{a_\bullet} \in (D^\times)^{*=-1}$ such that the following condition is satisfied:[3]

Let $G_{a_\bullet}$ be the algebraic group over $\mathbb{Q}$ such that $G_{a_\bullet}(R)$ for a $\mathbb{Q}$-algebra R consists of elements $g \in (D^{\mathrm{opp}} \otimes_{\mathbb{Q}} R)^\times$ with $g\beta_{a_\bullet} g^* = c(g)\beta_{a_\bullet}$ for some $c(g) \in R^\times$. If $G^1_{a_\bullet}$ denotes the kernel of the similitude character $c : G_{a_\bullet} \to \mathbb{G}_{m,\mathbb{Q}}$, then there exists an isomorphism

$$G^1_{a_\bullet}(\mathbb{R}) \simeq \prod_{i=1}^{f} U(a_i, n - a_i),$$

where the i-th factor corresponds to the real embedding $\tau_i : F \hookrightarrow \mathbb{R}$.

Note that the assumption on D at p implies that

$$G_{a_\bullet}(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_{\mathfrak{p}}) \cong \mathbb{Q}_p^\times \times \mathrm{GL}_n(\mathbb{Q}_{p^f}).$$

We put $V_{a_\bullet} = D$ and view it as a left D-module. Let $\langle -, - \rangle_{a_\bullet} : V_{a_\bullet} \times V_{a_\bullet} \to \mathbb{Q}$ be the perfect alternating pairing given by

$$\langle x, y \rangle_{a_\bullet} = \mathrm{Tr}_{D/\mathbb{Q}}(x\beta_{a_\bullet} y^*) \quad \text{for } x, y \in V_{a_\bullet}.$$

Then $G_{a_\bullet}$ is identified with the similitude group associated to $(V_{a_\bullet}, \langle -, - \rangle_{a_\bullet})$, i.e., for all $\mathbb{Q}$-algebra R, we have

$$G_{a_\bullet}(R) = \{g \in \mathrm{End}_{D\otimes_{\mathbb{Q}} R}(V_{a_\bullet} \otimes_{\mathbb{Q}} R) \mid \langle gx, gy \rangle_{a_\bullet} = c(g)\langle x, y \rangle_{a_\bullet} \text{ for some } c(g) \in R^\times\}.$$

Consider the homomorphism of $\mathbb{R}$-algebraic groups $h : \mathrm{Res}_{\mathbb{C}/\mathbb{R}}(\mathbb{G}_m) \to G_{a_\bullet,\mathbb{R}}$ given by

$$h(z) = \prod_{i=1}^{f} \mathrm{Diag}(\underbrace{z, \dots, z}_{a_i}, \underbrace{\bar{z}, \dots, \bar{z}}_{n-a_i}), \quad \text{for } z = x + \sqrt{-1}y. \tag{2.2.1}$$

Let $\mu_h : \mathbb{G}_{m,\mathbb{C}} \to G_{a_\bullet,\mathbb{C}}$ be the composite of $h_{\mathbb{C}}$ with the map

$$\mathbb{G}_{m,\mathbb{C}} \to \mathrm{Res}_{\mathbb{C}/\mathbb{R}}(\mathbb{G}_m)_{\mathbb{C}} \cong \mathbb{C}^\times \times \mathbb{C}^\times, \quad z \mapsto (z, 1).$$

Here, the first copy of $\mathbb{C}^\times$ in $\mathrm{Res}_{\mathbb{C}/\mathbb{R}}(\mathbb{G}_m)_{\mathbb{C}}$ is the one indexed by the identity element in $\mathrm{Aut}_{\mathbb{R}}(\mathbb{C})$, and the other copy of $\mathbb{C}^\times$ is indexed by the complex conjugation.

Let E_h be the reflex field of μ_h, i.e., the minimal subfield of $\mathbb{C}$ where the conjugacy class of μ_h is defined. It has the following explicit description. The group $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{C})$ acts naturally on $\Sigma_{\infty,E}$, and hence on the functions on $\Sigma_{\infty,E}$. Then E_h is the subfield of $\mathbb{C}$ fixed by the stabilizer of the $\mathbb{Z}$-valued function a on $\Sigma_{\infty,E}$ defined by $a(q_i) = a_i$ and $a(\bar{q}_i) = n - a_i$. The isomorphism $\iota_p : \mathbb{C} \xrightarrow{\sim} \overline{\mathbb{Q}}_p$

[3] As explained in the proof of [Harris and Taylor 2001, Lemma I.7.1], when n is odd, such $\beta_{a_\bullet}$ always exists, and when n is even, existence of $\beta_{a_\bullet}$ depends on the parity of $a_1 + \cdots + a_f$. See also the proof of Lemma 2.9.

defines a p-adic place $\wp$ of E_h. By our hypothesis on E, the local field $E_{h,\wp}$ is an unramified extension of $\mathbb{Q}_p$ contained in $\mathbb{Q}_{p^f}$, the unique unramified extension over $\mathbb{Q}_p$ of degree f.

2.3. ***Unitary Shimura varieties of PEL-type.*** Let $\mathcal{O}_D$ be a $*$-stable order of D and $\Lambda_{a_\bullet}$ an $\mathcal{O}_D$-lattice of $V_{a_\bullet}$ such that $\langle \Lambda_{a_\bullet}, \Lambda_{a_\bullet}\rangle_{a_\bullet} \subseteq \mathbb{Z}$ and $\Lambda_{a_\bullet} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is self-dual under the alternating pairing induced by $\langle -, -\rangle_{a_\bullet}$. We put $K_p = \mathbb{Z}_p^\times \times \mathrm{GL}_n(\mathcal{O}_{E_\mathfrak{p}}) \subseteq G_{a_\bullet}(\mathbb{Q}_p)$, and fix an open compact subgroup $K^p \subseteq G_{a_\bullet}(\mathbb{A}^{\infty,p})$ such that $K = K^p K_p$ is *neat*, i.e., $G_{a_\bullet}(\mathbb{Q}) \cap gKg^{-1}$ is torsion free for any $g \in G_{a_\bullet}(\mathbb{A}^\infty)$.

Following [Kottwitz 1992b], we have a unitary Shimura variety $\mathcal{S}h_{a_\bullet}$ defined over $\mathbb{Z}_{p^f}$;[4] it represents the functor that takes a locally noetherian $\mathbb{Z}_{p^f}$-scheme S to the set of isomorphism classes of tuples (A, λ, η), where

(1) A is an fn^2-dimensional abelian variety over S equipped with an action of $\mathcal{O}_D$ such that the induced action on $\mathrm{Lie}(A/S)$ satisfies the *Kottwitz determinant condition*, that is, if we view the *reduced* relative de Rham homology $H_1^{\mathrm{dR}}(A/S)^\circ := \mathfrak{e} H_1^{\mathrm{dR}}(A/S)$ and its quotient $\mathrm{Lie}^\circ_{A/S} := \mathfrak{e}\cdot \mathrm{Lie}_{A/S}$ as a module over $F_p \otimes_{\mathbb{Z}_p} \mathcal{O}_S \cong \bigoplus_{i=1}^f \mathcal{O}_S$, they, respectively, decompose into the direct sums of locally free $\mathcal{O}_S$-modules $H_1^{\mathrm{dR}}(A/S)_i^\circ$ of rank n and, their quotients, locally free $\mathcal{O}_S$-modules $\mathrm{Lie}^\circ_{A/S,i}$ of rank $n - a_i$;

(2) $\lambda : A \to A^\vee$ is a prime-to-p $\mathcal{O}_D$-equivariant polarization such that the Rosati involution induces the involution $*$ on $\mathcal{O}_D$;

(3) η is a collection of, for each connected component S_j of S with a geometric point $\bar{s}_j$, a $\pi_1(S_j, \bar{s}_j)$-invariant K^p-orbit of isomorphisms $\eta_j : \Lambda_{a_\bullet} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}^{(p)} \simeq T^{(p)}(A_{\bar{s}_j})$ such that the following diagram commutes for an isomorphism $\nu(\eta_j) \in \mathrm{Hom}(\widehat{\mathbb{Z}}^{(p)}, \widehat{\mathbb{Z}}^{(p)}(1))$:

$$\begin{tikzcd}
\Lambda_{a_\bullet} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}^{(p)} \times \Lambda_{a_\bullet} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}^{(p)} \arrow[r, "\langle -,- \rangle"] \arrow[d, "\eta_j \times \eta_j"] & \widehat{\mathbb{Z}}^{(p)} \arrow[d, "\nu(\eta_j)"] \\
T^{(p)} A_{\bar{s}_j} \times T^{(p)} A_{\bar{s}_j} \arrow[r, "\text{Weil pairing}"] & \widehat{\mathbb{Z}}^{(p)}(1),
\end{tikzcd}$$

where $\widehat{\mathbb{Z}}^{(p)} = \prod_{\ell \neq p} \mathbb{Z}_\ell$ and $T^{(p)}(A_{\bar{s}_j})$ denotes the product of the ℓ-adic Tate modules of $A_{\bar{s}_j}$ for all $\ell \neq p$.

The Shimura variety $\mathcal{S}h_{a_\bullet}$ is smooth and projective over $\mathbb{Z}_{p^f}$ of relative dimension $d(a_\bullet) := \sum_{i=1}^f a_i(n - a_i)$. Note that if $a_i \in \{0, n\}$ for all i, then $\mathcal{S}h_{a_\bullet}$ is of relative dimension zero; we call it a *discrete Shimura variety*.

[4] Although one can descend $\mathcal{S}h_{a_\bullet}$ to the subring $\mathcal{O}_{E_{h,\wp}}$ of $\mathbb{Z}_{p^f}$, we ignore this minor improvement.

We denote by $\mathcal{S}h_{a_\bullet}(\mathbb{C})$ the complex points of $\mathcal{S}h_{a_\bullet}$ via the embedding

$$\mathbb{Z}_{p^f} \hookrightarrow \overline{\mathbb{Q}}_p \xrightarrow{\iota_p^{-1}} \mathbb{C}.$$

Let $K_\infty \subseteq G_{a_\bullet}(\mathbb{R})$ be the stabilizer of h (2.2.1) under the conjugation action, and let X_∞ denote the $G_{a_\bullet}(\mathbb{R})$-conjugacy class of h. Then K_∞ is a maximal compact-modulo-center subgroup of $G_{a_\bullet}(\mathbb{R})$. According to [Kottwitz 1992b, page 400], the complex manifold $\mathcal{S}h_{a_\bullet}(\mathbb{C})$ is the disjoint union of $\#\ker^1(\mathbb{Q}, G_{a_\bullet})$ copies of

$$G_{a_\bullet}(\mathbb{Q}) \backslash (G_{a_\bullet}(\mathbb{A}^\infty) \times X_\infty)/K \cong G_{a_\bullet}(\mathbb{Q}) \backslash G_{a_\bullet}(\mathbb{A})/K \times K_\infty. \tag{2.3.1}$$

Here, if n is even, then $\ker^1(\mathbb{Q}, G_{a_\bullet}) = (0)$, while if n is odd then

$$\ker^1(\mathbb{Q}, G_{a_\bullet}) = \mathrm{Ker}(F^\times/\mathbb{Q}^\times N_{E/F}(E^\times) \to \mathbb{A}_F^\times/\mathbb{A}^\times N_{E/F}(\mathbb{A}_E^\times)).$$

In either case, $\ker^1(\mathbb{Q}, G_{a_\bullet})$ depends only on the CM extension E/F and the parity of n but not on the tuple $a_\bullet$.

Let $\mathrm{Sh}_{a_\bullet} := \mathcal{S}h_{a_\bullet} \otimes_{\mathbb{Z}_{p^f}} \mathbb{F}_{p^f}$ denote the special fiber of $\mathcal{S}h_{a_\bullet}$, and let $\overline{\mathrm{Sh}}_{a_\bullet} := \mathrm{Sh}_{a_\bullet} \otimes_{\mathbb{F}_{p^f}} \bar{\mathbb{F}}_p$ denote the geometric special fiber.

2.4. *ℓ-adic cohomology.* We fix a prime number $\ell \neq p$, and an isomorphism $\iota_\ell : \mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$. Let ξ be an algebraic representation of $G_{a_\bullet}$ over $\overline{\mathbb{Q}}_\ell$, and $\xi_\mathbb{C}$ be the base change via ι_ℓ^{-1}. The theory of automorphic sheaves [Milne 1990, Section III] or just reading off from the rational ℓ-adic Tate modules of the universal abelian variety allows us to attach to ξ a lisse $\overline{\mathbb{Q}}_\ell$-sheaf $\mathcal{L}_\xi$ over $\mathcal{S}h_{a_\bullet}$. For example, if ξ is the representation of $G_{a_\bullet}$ on the vector space $V_{a_\bullet}$ (Section 2.2), the corresponding ℓ-adic local system is given by the rational ℓ-adic Tate module (tensored with $\overline{\mathbb{Q}}_\ell$) of the universal abelian scheme over $\mathcal{S}h_{a_\bullet}$.

We assume that ξ is irreducible. Let $\mathscr{H}_K = \mathscr{H}(K, \overline{\mathbb{Q}}_\ell)$ be the Hecke algebra of compactly supported K-bi-invariant $\overline{\mathbb{Q}}_\ell$-valued functions on $G_{a_\bullet}(\mathbb{A}^\infty)$. The étale cohomology group $H^{d(a_\bullet)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{a_\bullet}, \mathcal{L}_\xi)$ is equipped with a natural action of $\mathscr{H}_K \times \mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^f})$. Since $\mathrm{Sh}_{a_\bullet}$ is proper and smooth, there is no continuous spectrum and we have a canonical decomposition of $\mathscr{H}_K \times \mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^f})$-modules (see, e.g., [Harris and Taylor 2001, Proposition III.2.1])

$$H^{d(a_\bullet)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{a_\bullet}, \mathcal{L}_\xi) = \bigoplus_{\pi \in \mathrm{Irr}(G_{a_\bullet}(\mathbb{A}^\infty))} \iota_\ell(\pi^K) \otimes R_{a_\bullet,\ell}(\pi), \tag{2.4.1}$$

where $\mathrm{Irr}(G_{a_\bullet}(\mathbb{A}^\infty))$ is the set of irreducible admissible representations of $G_{a_\bullet}(\mathbb{A}^\infty)$ with coefficients in $\mathbb{C}$, π^K is the K-invariant subspace of $\pi \in \mathrm{Irr}(G_{a_\bullet}(\mathbb{A}^\infty))$ and $R_{a_\bullet,\ell}(\pi)$ is a certain ℓ-adic representation of $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^f})$ which we specify below.

We write $H^{d(a_\bullet)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{a_\bullet}, \mathcal{L}_\xi)_\pi$ for the *π-isotypic component* of the cohomology group, that is, the direct summand of (2.4.1) labeled by π. We make the following assumptions on π.

Hypothesis 2.5. (1) We have $\pi^K \neq 0$.

(2) There exists an admissible irreducible representation π_∞ of $G_{a_\bullet}(\mathbb{R})$ such that $\pi \otimes \pi_\infty$ is a cuspidal automorphic representation of $G_{a_\bullet}(\mathbb{A})$,

(2a) π_∞ is *cohomological in degree $d(a_\bullet)$ for ξ* in the sense that

$$H^{d(a_\bullet)}(\mathrm{Lie}(G_{a_\bullet}(\mathbb{R})), K_\infty, \pi_\infty \otimes \xi_{\mathbb{C}}) \neq 0, \tag{2.5.1}$$ [5]

where K_∞ is a maximal compact subgroup of $G_{a_\bullet}(\mathbb{R})$,

(2b) and $\pi \otimes \pi_\infty$ admits a base change to a *cuspidal* automorphic representation of $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$.

Note that Hypothesis 2.5(1) implies that the p-component π_p is unramified. Hypothesis 2.5 (2a) ensures that $R_{a_\bullet,\ell}(\pi)$ is nontrivial. Moreover, by [Caraiani 2012, Theorem 1.2], this hypothesis implies that the base change of $\pi \otimes \pi_\infty$ to $\mathrm{GL}_{n,E}$ is tempered at all finite places, and hence π_p is tempered.

We recall now an explicit description, due to Kottwitz [1992a], of the Galois module $R_{a_\bullet,\ell}(\pi)$. As $G_{a_\bullet}(\mathbb{Q}_p) = \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_{\mathfrak{p}})$, we may write $\pi_p = \pi_{p,0} \otimes \pi_{\mathfrak{p}}$, where $\pi_{p,0}$ is a character of $\mathbb{Q}_p^\times$ trivial on $\mathbb{Z}_p^\times$, and $\pi_{\mathfrak{p}}$ is an irreducible admissible representation of $\mathrm{GL}_n(E_{\mathfrak{p}})$ such that $\pi_{\mathfrak{p}}^{\mathrm{GL}_n(\mathcal{O}_{E_{\mathfrak{p}}})} \neq 0$. Choose a square root $\sqrt{p}$ of p in $\overline{\mathbb{Q}}$. Depending on this choice of $\sqrt{p}$, one has an (unramified) local Langlands parameter attached to π_p:

$$\varphi_{\pi_p} = (\varphi_{\pi_{p,0}}, \varphi_{\pi_{\mathfrak{p}}}) : W_{\mathbb{Q}_p} \to {}^L(G_{a_\bullet,\mathbb{Q}_p}) \simeq \mathbb{C}^\times \times (\mathrm{GL}_n(\mathbb{C})^{\mathbb{Z}/f\mathbb{Z}} \rtimes \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)).$$

Here, $W_{\mathbb{Q}_p}$ is the Weil group of $\mathbb{Q}_p$, and $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ permutes cyclically the f copies of $\mathrm{GL}_n(\mathbb{C})$ though the quotient $\mathrm{Gal}(\mathbb{Q}_{p^f}/\mathbb{Q}_p) \cong \mathbb{Z}/f\mathbb{Z}$. The image of $\varphi_{\pi_p}|_{W_{\mathbb{Q}_{p^f}}}$ lies in $({}^L G_{a_\bullet})^\circ \simeq \mathbb{C}^\times \times \mathrm{GL}_n(\mathbb{C})^{\mathbb{Z}/f\mathbb{Z}}$. The cocharacter $\mu_h : \mathbb{G}_{m,E_h} \to G_{a_\bullet,E_h}$ induces a character $\check{\mu}_h$ of $({}^L G_{a_\bullet})^\circ$ over E_h. Let r_{μ_h} denote the algebraic representation of $({}^L G_{a_\bullet})^\circ$ with extreme weight $\check{\mu}_h$. Denote by Frob_{p^f} a *geometric* Frobenius element in $W_{\mathbb{Q}_{p^f}}$. Let $\overline{\mathbb{Q}}_\ell(1/2)$ denote the unramified representation of $W_{\mathbb{Q}_{p^f}}$ which sends Frob_{p^f} to $(\sqrt{p})^{-f}$. Then $R_{a_\bullet,\ell}(\pi)$ can be described in terms of φ_{π_p} as follows.

Theorem 2.6 [Kottwitz 1992a, Theorem 1]. *Under the hypothesis and notation above, we have an equality in the Grothendieck group of $W_{\mathbb{Q}_{p^f}}$-modules*:

$$[R_{a_\bullet,\ell}(\pi)] = \#\ker^1(\mathbb{Q}, G_{a_\bullet}) m_{a_\bullet}(\pi) \big[\iota_\ell(r_{\mu_h} \circ \varphi_{\pi_p}) \otimes \overline{\mathbb{Q}}_\ell\big(-\tfrac{1}{2} d(a_\bullet)\big)\big],$$

where $m_{a_\bullet}(\pi)$ is a certain integer related to the automorphic multiplicities of automorphic representations of $G_{a_\bullet}$ with finite part π. [6]

[5] This automatically implies that π_∞ has the same central and infinitesimal characters as the *contragradient* of $\xi_{\mathbb{C}}$.

[6] Rigorously speaking, Kottwitz's theorem describes the direct sum of the π-component of all cohomological degrees. Since our π_p is tempered, so π appears only in the middle degree for purity reasons because $\mathrm{Sh}_{a_\bullet}$ is compact.

In our case, one can make Kottwitz's theorem more transparent. Define an ℓ-adic representation

$$\rho_{\pi_\mathfrak{p}} = \iota_\ell(\varphi_{\pi_\mathfrak{p}}^{(1),\vee}) \otimes \overline{\mathbb{Q}}_\ell\big(\tfrac{1}{2}(1-n)\big) : W_{\mathbb{Q}_{p^f}} \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell), \tag{2.6.1}$$

where $\varphi_{\pi_\mathfrak{p}}^{(1),\vee} : W_{\mathbb{Q}_{p^f}} \to \mathrm{GL}_n(\mathbb{C})$ denotes the *contragredient* of the projection to the first (or any) copy of $\mathrm{GL}_n(\mathbb{C})$. Both $\varphi_{\pi_\mathfrak{p}}$ and $\overline{\mathbb{Q}}_\ell\big(\tfrac{1}{2}\big)$ depend on the choice of $\sqrt{p}$, but $\rho_{\pi_\mathfrak{p}}$ does not. Explicitly, $\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})$ is semisimple with the characteristic polynomial given by [Gross 1998, (6.7)]:

$$X^n + \sum_{i=1}^{n} (-1)^i (N\mathfrak{p})^{i(i-1)/2} a_\mathfrak{p}^{(i)} X^{n-i}, \tag{2.6.2}$$

where $a_\mathfrak{p}^{(i)}$ is the eigenvalue on $\pi_\mathfrak{p}^{\mathrm{GL}_n(\mathcal{O}_{E_\mathfrak{p}})}$ of the Hecke operator

$$T_\mathfrak{p}^{(i)} = \mathrm{GL}_n(\mathcal{O}_{E_\mathfrak{p}}) \cdot \mathrm{Diag}(\underbrace{p, \dots, p}_{i}, \underbrace{1, \dots, 1}_{n-i}) \cdot \mathrm{GL}_n(\mathcal{O}_{E_\mathfrak{p}}).$$

An easy computation shows that $r_{\mu_h} = \mathrm{Std}_{\mathbb{Q}_p^\times}^{-1} \otimes \bigotimes_{i=1}^{f} (\wedge^{a_i} \mathrm{Std}^\vee)$. Since the projection of $\varphi_{\pi_p}|_{W_{\mathbb{Q}_{p^f}}}$ to each copy of $\mathrm{GL}_n(\mathbb{C})$ is conjugate to all others, Theorem 2.6 is equivalent to

$$\begin{aligned}[][R_{a_\bullet,\ell}(\pi)] \\ &= \#\ker^1(\mathbb{Q}, G_{a_\bullet}) \cdot m_{a_\bullet}(\pi) \Big[\rho_{a_\bullet}(\pi_\mathfrak{p}) \otimes \chi_{\pi_{p,0}}^{-1} \otimes \overline{\mathbb{Q}}_\ell\Big(\sum_i \tfrac{1}{2} a_i(a_i - 1)\Big)\Big],\end{aligned} \tag{2.6.3}$$

where $\rho_{a_\bullet}(\pi_\mathfrak{p}) = r_{a_\bullet} \circ \rho_{\pi_\mathfrak{p}}$ with $r_{a_\bullet} = \bigotimes_{i=1}^{f} \wedge^{a_i} \mathrm{Std}$, and $\chi_{\pi_{p,0}}$ denotes the character of $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_{p^f})$ sending Frob_{p^f} to $\iota_\ell(\pi_{p,0}(p^f))$.

Remark 2.7. The reason why we normalize the Galois representation as above is the following: By Hypothesis 2.5, π is the finite part of an automorphic representation of $G_{a_\bullet}(\mathbb{A})$ which admits a base change to a cuspidal automorphic representation $\Pi \otimes \chi$ of $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$. If ρ_Π denotes the Galois representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/E)$ attached to Π, then $\rho_{\pi_\mathfrak{p}}$ is the semisimplification of the restriction of ρ_Π to $W_{E_\mathfrak{p}}$ (See [Caraiani 2012, Theorem 1.1]).

2.8. ***Tate conjecture.*** We recall first the Tate conjecture [1966] over finite fields. Let X be a projective smooth variety over a finite field $\mathbb{F}_q$ of characteristic p. Put $\overline{X} = X_{\overline{\mathbb{F}}_p}$. For each prime $\ell \neq p$ and integer $r \leq \dim(X)$, we have a cycle class map

$$\mathrm{cl}_X^r : A^r(X) \otimes_\mathbb{Z} \overline{\mathbb{Q}}_\ell \to H^{2r}_{\mathrm{et}}(\overline{X}, \overline{\mathbb{Q}}_\ell(r))^{\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_q)},$$

where $A^r(X)$ denotes the abelian group of codimension r algebraic cycles in X defined over $\mathbb{F}_q$. Then the Tate conjecture predicts that this map is surjective. One has a geometric variant of the Tate conjecture, which claims that the geometric cycle class map:

$$\mathrm{cl}^r_{\bar{X}} : A^r(\bar{X}) \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}_\ell \to H^{2r}_{\mathrm{et}}(\bar{X}, \overline{\mathbb{Q}}_\ell(r))^{\mathrm{fin}} := \bigcup_{m \geq 1} H^{2r}_{\mathrm{et}}(\bar{X}, \overline{\mathbb{Q}}_\ell(r))^{\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{q^m})}$$

is surjective. Here, the superscript "fin" means the subspace on which $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_q)$ acts through a finite quotient. Note that the surjectivity of $\mathrm{cl}^r_{\bar{X}}$ implies that of cl^r_X by taking the $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_q)$-invariant subspace.

Consider the case $X = \mathrm{Sh}_{a_\bullet}$ with $d(a_\bullet)$ even. Let π be an irreducible admissible representation of $G_{a_\bullet}(\mathbb{A}^\infty)$ as in Theorem 2.6. By Theorem 2.6, the π-isotypic component of $H^{d(a_\bullet)}_{\mathrm{et}}\big(\overline{\mathrm{Sh}}_{a_\bullet}, \overline{\mathbb{Q}}_\ell\big(\tfrac{1}{2}d(a_\bullet)\big)\big)^{\mathrm{fin}}$ is, up to Frobenius semisimplification[7], isomorphic to $\dim(\pi^K) \cdot \#\ker^1(\mathbb{Q}, G_{a_\bullet}) \cdot m_{a_\bullet}(\pi)$ copies of

$$\left(\rho_{a_\bullet}(\pi_{\mathfrak{p}}) \otimes \chi^{-1}_{\pi_{p,0}} \otimes \overline{\mathbb{Q}}_\ell\left(\frac{(n-1)}{2}\sum_{i=1}^{f} a_i\right)\right)^{\mathrm{fin}}. \tag{2.8.1}$$

Note that $\chi_{\pi_{p,0}}(\mathrm{Frob}_{p^f}) = \pi_{p,0}(p^f)$ is a root of unity. Hence, the dimension of (2.8.1) is equal to the sum of the dimensions of the Frob_{p^f}-eigenspaces of $\rho_{a_\bullet}(\pi_{\mathfrak{p}})$ with eigenvalues $(p^f)^{(n-1)/2\sum_i a_i}\zeta$ for some root of unity ζ. In many examples, this space is known to be nonzero.

For instance, when $f = 2$, $a_1 = r$ and $a_2 = n - r$ for some $1 \leq r \leq n-1$, we have $d(a_\bullet) = 2r(n-r)$ and

$$\rho_{a_\bullet}(\pi_{\mathfrak{p}}) = \wedge^r \rho_{\pi_{\mathfrak{p}}} \otimes \wedge^{n-r} \rho_{\pi_{\mathfrak{p}}}.$$

Let $V_{\pi_{\mathfrak{p}}, a_\bullet}$ denote the space of representation $\rho_{a_\bullet}(\pi_{\mathfrak{p}})$. If $\rho_{\pi_{\mathfrak{p}}}(\mathrm{Frob}_{p^f})$ has distinct eigenvalues $\alpha_1, \dots, \alpha_n$, then the eigenvalues of Frob_{p^f} on $V_{\pi_{\mathfrak{p}}, a_\bullet}$ are given by $\alpha_{i_1} \cdots \alpha_{i_r} \cdot \alpha_{j_1} \cdots \alpha_{j_{n-r}}$, for distinct subscripts $i_1, \dots, i_r$ and distinct subscripts $j_1, \dots, j_{n-r}$. This product is exactly $(p^f)^{n(n-1)/2} a^{(n)}_{\mathfrak{p}}$ (note that $a^{(n)}_{\mathfrak{p}}$ is a root of unity) if the set $\{i_1, \dots, i_r\}$ and the set $\{j_1, \dots, j_{n-r}\}$ are the complement of each other as subsets of $\{1, \dots, n\}$. On the other hand, if the subsets $\{i_1, \dots, i_r\}$ and $\{j_1, \dots, j_{n-r}\}$ are not the complement of each other and if the α_i are "sufficiently generic"[8], the eigenvalue $\alpha_{i_1} \cdots \alpha_{i_r} \cdot \alpha_{j_1} \cdots \alpha_{j_{n-r}}$ is not a root of unity. In other words, the dimension of (2.8.1) is "generically" equal to $\binom{n}{r}$. As predicted by the Tate conjecture, these cohomology classes should come from algebraic cycles. Our main conjecture addresses exactly this, and it predicts that those desired "generic"

[7] Conjecturally, the Frobenius action on the étale ℓ-adic cohomology groups of a projective smooth variety over a finite field is always semisimple.

[8] For example, if $r = 1$ and $\alpha_1 = \alpha_2$, the eigenvalues $\alpha_1 \cdot \alpha_1\alpha_3\alpha_4 \cdots \alpha_n$ is equal to $\alpha_1 \cdots \alpha_n$ and hence is $p^{n(n-1)}$ times a root of unity. So to be in the generic case, we will need to require that α_i/α_j for $i \neq j$ is not a root of unity if $r = 1$. For another example, if $r = 2$, "generic" will mean that α_i/α_j for $i \neq j$ and $\alpha_i\alpha_{i'}/\alpha_j\alpha_{j'}$ for $\{i, i'\} \neq \{j, j'\}$ are not roots of unity.

algebraic cycles can be given by the irreducible components of the basic locus, and are birationally equivalent to certain fiber bundles over the special fiber of some other Shimura varieties associated to inner forms of $G_{a_\bullet}$. To make this precise, we need the following lemma.

Lemma 2.9. *Let $b_\bullet = (b_i)_{1 \le i \le f}$ be a tuple with $b_i \in \{0, \dots, n\}$ such that $\sum_{i=1}^{f} b_i \equiv \sum_{i=1}^{f} a_i \pmod 2$ if n is even. Then there exists $\beta_{b_\bullet} \in (D^\times)^{*=-1}$ such that*

- *the alternating D-Hermitian space $(V_{b_\bullet}, \langle -, - \rangle_{b_\bullet})$ defined using $\beta_{b_\bullet}$ in place of $\beta_{a_\bullet}$ is isomorphic to $(V_{a_\bullet}, \langle -, - \rangle_{a_\bullet})$ when tensored with $\mathbb{A}^\infty$, and*
- *if $G_{b_\bullet}$ denotes the corresponding algebraic group over $\mathbb{Q}$ defined in the similar way with $\beta_{a_\bullet}$ replaced by $\beta_{b_\bullet}$, then*

$$G^1_{b_\bullet}(\mathbb{R}) \simeq \prod_{i=1}^{f} U(b_i, n - b_i).$$

Proof. We reduce the problem to the existence of a certain cohomology class. Note that $G^1_{a_\bullet} = \mathrm{Aut}(V_{a_\bullet}, \langle -, - \rangle_{a_\bullet})$ is the Weil restriction to $\mathbb{Q}$ of a unitary group $U_{a_\bullet}$ over F. The cohomology set $H^1(\mathbb{Q}, G^1_{a_\bullet}) \cong H^1(F, U_{a_\bullet})$ is in bijection with the isomorphism classes of one-dimensional skew-Hermitian D-modules V. As $U_{a_\bullet} \times_F E \simeq \mathrm{GL}_{n,E}$, Hilbert's Theorem 90 for GL_n implies that the inflation map induces an isomorphism

$$H^1(E/F, U_{a_\bullet}) \xrightarrow{\sim} H^1(F, U_{a_\bullet}).$$

Denote by $g \mapsto g^{\sharp_{\beta_{a_\bullet}}} = \beta_{a_\bullet} g^* \beta_{a_\bullet}^{-1}$ the involution on D induced by the alternating pairing $\langle -, - \rangle_{a_\bullet}$. Then a 1-cocycle of $\mathrm{Gal}(E/F)$ with values in $U_{a_\bullet}$ is given by an element $\alpha \in D^\times$ such that $\alpha = \alpha^{\sharp_{\beta_{a_\bullet}}}$, and $\alpha_1, \alpha_2 \in D^\times$ define the same cohomology class in $H^1(F, U_{a_\bullet})$ if and only if there exists $g \in D^\times$ such that $g\alpha_1 g^{\sharp_{\beta_{a_\bullet}}} = \alpha_2$. Explicitly, given such an α, the corresponding skew-Hermitian D-module is given by $V = D$ equipped with the alternating pairing

$$\langle -, - \rangle_\alpha : V \times V \to \mathbb{Q}, \qquad (x, y) \mapsto \mathrm{Tr}_{D/\mathbb{Q}}(x\alpha\beta_{a_\bullet} y^*).$$

For a place v of F, we denote by

$$\mathrm{loc}_v : H^1(F, U_{a_\bullet}) \to H^1(F_v, U_{a_\bullet})$$

the canonical localization map. By [Helm 2012, Proposition 8.1], if $\sum_{i=1}^{f} b_i \equiv \sum_{i=1}^{f} a_i \mod 2$, there exists a cohomology class $[\alpha] \in H^1(F, U_{a_\bullet})$ such that

- $\mathrm{loc}_v([\alpha])$ is trivial for every finite place v of F, and
- if $v = \tau_i$ with $1 \le i \le n$ is an archimedean place, one has an isomorphism of unitary groups over $\mathbb{R}$: $\mathrm{Aut}(V \otimes_{F,\tau_i} \mathbb{R}, \langle -, - \rangle_\alpha) \simeq U(b_i, n - b_i)$.

Then the element $\beta_{b_\bullet} = \alpha\beta_{a_\bullet}$ meets the requirements of Lemma 2.9. $\square$

In the sequel, we always fix a choice of $\beta_{b_\bullet}$, and as well as an isomorphism $\gamma_{a_\bullet,b_\bullet} : V_{a_\bullet} \otimes_{\mathbb{Q}} \mathbb{A}^\infty \xrightarrow{\sim} V_{b_\bullet} \otimes_{\mathbb{Q}} \mathbb{A}^\infty$, which induces an isomorphism $G_{a_\bullet}(\mathbb{A}^\infty) \simeq G_{b_\bullet}(\mathbb{A}^\infty)$. Recall that we have chosen a lattice $\Lambda_{a_\bullet} \subseteq V_{a_\bullet}$ to define the moduli problem for $\mathcal{S}h_{a_\bullet}$. We put $\Lambda_{b_\bullet} := V_{b_\bullet} \cap \gamma_{a_\bullet,b_\bullet}(\Lambda_{a_\bullet} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}})$. Then applying the construction of Section 2.3 to the lattice $\Lambda_{b_\bullet} \subseteq V_{b_\bullet}$ and the open compact subgroup $K^p \subseteq G_{a_\bullet}(\mathbb{A}^{\infty,p}) \simeq G_{b_\bullet}(\mathbb{A}^{\infty,p})$, we get a Shimura variety $\mathcal{S}h_{b_\bullet}$ over $\mathbb{Z}_{p^f}$ of level K^p as well as its special fiber $\mathrm{Sh}_{b_\bullet}$. Moreover, an algebraic representation ξ of $G_{a_\bullet}$ over $\overline{\mathbb{Q}}_\ell$ corresponds, via the fixed isomorphism $G_{a_\bullet}(\mathbb{A}^\infty) \simeq G_{b_\bullet}(\mathbb{A}^\infty)$, to an algebraic representation of $G_{b_\bullet}$ over $\overline{\mathbb{Q}}_\ell$. We use the same notation $\mathcal{L}_\xi$ to denote the étale sheaf on $\mathcal{S}h_{a_\bullet}$ and $\mathcal{S}h_{b_\bullet}$ defined by ξ.

2.10. ***Gysin/trace maps.*** Before stating the main conjecture of this paper, we recall the general definition of Gysin maps. Let $f : Y \to X$ be a proper morphism of smooth varieties over an algebraically closed field k. Let d_X and d_Y be the dimensions of X and Y respectively. Recall that the derived direct image Rf_* on the derived category of constructible ℓ-adic étale sheaves has a left adjoint $f^!$. Since both X and Y are smooth, the ℓ-adic dualizing complex of X (resp. Y) is $\overline{\mathbb{Q}}_\ell(d_X)[2d_X]$ (resp. $\overline{\mathbb{Q}}_\ell(d_Y)[2d_Y]$). Therefore, one has

$$f^!(\overline{\mathbb{Q}}_\ell(d_X)[2d_X]) = \overline{\mathbb{Q}}_\ell(d_Y)[2d_Y].$$

The adjunction map $Rf_*f^!\overline{\mathbb{Q}}_\ell \to \overline{\mathbb{Q}}_\ell$ induces a canonical morphism

$$\mathrm{Tr}_f : Rf_*\overline{\mathbb{Q}}_\ell \to \overline{\mathbb{Q}}_\ell(d_X - d_Y)[2(d_X - d_Y)].$$

More generally, if $\mathcal{L}$ is a lisse $\overline{\mathbb{Q}}_\ell$-sheaf on X, it induces a Gysin/trace map

$$Rf_*(f^*\mathcal{L}) \cong \mathcal{L} \otimes Rf_*(\overline{\mathbb{Q}}_\ell) \xrightarrow{1\otimes \mathrm{Tr}_f} \mathcal{L}(d_X - d_Y)[2(d_X - d_Y)],$$

where the first isomorphism is the projection formula [SGA 4_2 1972, XVII 5.2.9]. When f is flat with equidimensional fibers of dimension $d_Y - d_X$, this is the trace map as defined in [SGA 4_2 1972, XVIII 2.9]. When f is a closed immersion of codimension $r = d_X - d_Y$, it is the usual Gysin map. For any integer q, the Gysin/trace map induces a morphism on cohomology groups:

$$f_! : H^q_{\mathrm{et}}(Y, f^*\mathcal{L}) \to H^{q+2(d_X-d_Y)}_{\mathrm{et}}(X, \mathcal{L}(d_X - d_Y)). \tag{2.10.1}$$

2.11. ***Representation theory of*** $\mathbf{GL}_n$. As suggested by the description of Galois representations appearing in the middle cohomology group of Shimura varieties in Theorem 2.6, as well as by the Tate conjecture, we need to understand the representation theory of GL_n embedded diagonally into the Langlands dual group

$$({}^L G_{a_\bullet})^\circ \simeq \mathbb{C}^\times \times \mathrm{GL}_n(\mathbb{C})^{\mathbb{Z}/f\mathbb{Z}}.$$

The Hodge cocharacter μ of $G_{a_\bullet}$ gives rise to the representation $r_{a_\bullet}=\bigotimes_{i=1}^{f}(\wedge^{a_i}\,\mathrm{Std})$ of the diagonal GL_n. If λ is a dominant weight of GL_n (with respect to the usual diagonal torus and upper triangular Borel subgroup) appearing in $r_{a_\bullet}$, we can write this weight λ as the sum of f dominant minuscule weights $\omega_{b_1}+\cdots+\omega_{b_f}$, where ω_i for $0\le i\le n$ is the weight of GL_n that takes $\mathrm{Diag}(t_1,\dots,t_n)$ to $t_1\cdots t_i$. The set $\{b_1,\dots,b_f\}$ (counted with multiplicity) is unique, which we denote by B_λ. Explicitly, if λ takes $\mathrm{Diag}(t_1,\dots,t_n)$ to $t_1^{\beta_1}\cdots t_n^{\beta_n}$ (necessarily $\beta_1\le f$), then

$$B_\lambda=\{\underbrace{n,\dots,n}_{\beta_n},\underbrace{n-1,\dots,n-1}_{\beta_{n-1}-\beta_n},\dots,\underbrace{1,\dots,1}_{\beta_1-\beta_0},\underbrace{0,\dots,0}_{f-\beta_1}\}.$$

Moreover, we always have $\sum a_i=\sum b_i$. In particular, this implies by Lemma 2.9 that the Shimura variety $\mathrm{Sh}_{b_\bullet}$ makes sense, and the étale sheaf $\mathcal{L}_\xi$ is well defined on $\mathrm{Sh}_{b_\bullet}$.

We write $m_\lambda(a_\bullet)$ for the multiplicity of the weight λ in $r_{a_\bullet}$.

Conjecture 2.12. *Let $\mathrm{Sh}_{a_\bullet}$ and $\mathcal{L}_\xi$ be as in Section 2.4. Let λ be a dominant weight that appears in the representation $r_{a_\bullet}$ as in Section 2.11. Define B_λ and $m_\lambda(a_\bullet)$ as in Section 2.11.*

Then there exist varieties $Y_1,\dots,Y_{m_\lambda(a_\bullet)}$ of dimension $\frac{1}{2}(d(a_\bullet)+d(b_\bullet))$ over $\mathbb{F}_{p^f}$, equipped with natural action of prime-to-p Hecke correspondences, such that each Y_j fits into a diagram

$$\begin{tikzcd} & Y_j \arrow[dl, "\mathrm{pr}_{a_\bullet}^{(j)}"'] \arrow[dr, "\mathrm{pr}_{b_\bullet^{(j)}}"] & \\ \mathrm{Sh}_{a_\bullet} & & \mathrm{Sh}_{b_\bullet^{(j)}} \end{tikzcd}$$

satisfying the following properties.

(1) *For each j, $b_\bullet^{(j)}=(b_1^{(j)},\dots,b_f^{(j)})$ is a reordering of the elements of the set B_λ, and both $\mathrm{pr}_{a_\bullet}^{(j)}$ and $\mathrm{pr}_{b_\bullet^{(j)}}$ are equivariant for the prime-to-p Hecke correspondences.*

(2) *The morphism $\mathrm{pr}_{a_\bullet}^{(j)}$ is a proper morphism and is birational onto the image. The morphism $\mathrm{pr}_{b_\bullet^{(j)}}$ is proper and generically smooth of relative dimension $\frac{1}{2}(d(a_\bullet)-d(b_\bullet))$ (note that $d(b_\bullet)\equiv d(a_\bullet)\pmod 2$ since $\sum_i a_i=\sum_i b_i$).*

(3) *There exists a p-isogeny of abelian schemes over Y_j*

$$\phi_{b_\bullet^{(j)},a_\bullet}:\mathrm{pr}^*_{b_\bullet^{(j)}}(\mathcal{A}_{b_\bullet^{(j)}})\to \mathrm{pr}_{a_\bullet}^{(j),*}(\mathcal{A}_{a_\bullet}),$$

where $\mathcal{A}_{a_\bullet}$ and $\mathcal{A}_{b_\bullet^{(j)}}$ denote respectively the universal abelian scheme on $\mathrm{Sh}_{a_\bullet}$ and $\mathrm{Sh}_{b_\bullet^{(j)}}$. Let

$$\phi_{b_\bullet^{(j)},a_\bullet,*}:\mathrm{pr}^*_{b_\bullet^{(j)}}\mathcal{L}_\xi\xrightarrow{\sim}\mathrm{pr}_{a_\bullet}^{(j),*}\mathcal{L}_\xi.$$

be the isomorphism of the ℓ-adic sheaves induced by $\phi_{b_\bullet^{(j)},a_\bullet}$ via the construction in Section 2.4.[9]

(4) *Let π be an irreducible admissible representation of $G_{a_\bullet}(\mathbb{A}^\infty)\simeq G_{b_\bullet^{(j)}}(\mathbb{A}^\infty)$ satisfying Hypothesis 2.5 for both $a_\bullet$ and $b_\bullet$, and assume that $m_{a_\bullet}(\pi)=m_{b_\bullet^{(j)}}(\pi)$ for all j* [10]. *Suppose that the n eigenvalues $\alpha_1,\dots,\alpha_n$ of $\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})$ are "sufficiently generic" in the sense that the generalized eigenspace decomposition of $\rho_{a_\bullet}(\mathrm{Frob}_{p^N})$ for any large N is the same as the weight space decomposition of the algebraic representation $r_{a_\bullet}$. Then the natural homomorphism of π-isotypic components*[11] *of the cohomology groups*

$$\bigoplus_{j=1}^{m_\lambda(a_\bullet)} H_{\mathrm{et}}^{d(b_\bullet)}\big(\overline{\mathrm{Sh}}_{b_\bullet^{(j)}},\mathcal{L}_\xi\big(\tfrac12 d(b_\bullet)\big)\big)_\pi^{\mathrm{Frob}_{p^f}=\lambda}$$

$$\xrightarrow{\oplus\,\mathrm{pr}^*_{b_\bullet^{(j)}}} \bigoplus_{j=1}^{m_\lambda(a_\bullet)} H_{\mathrm{et}}^{d(b_\bullet)}\big(\bar Y_j,\mathrm{pr}^*_{b_\bullet^{(j)}}\mathcal{L}_\xi\big(\tfrac12 d(b_\bullet)\big)\big)_\pi^{\mathrm{Frob}_{p^f}=\lambda}$$

$$\xrightarrow{\oplus\phi_{b_\bullet^{(j)},a_\bullet,*}} \bigoplus_{j=1}^{m_\lambda(a_\bullet)} H_{\mathrm{et}}^{d(b_\bullet)}\big(\bar Y_j,\mathrm{pr}^*_{a_\bullet}\mathcal{L}_\xi\big(\tfrac12 d(b_\bullet)\big)\big)_\pi^{\mathrm{Frob}_{p^f}=\lambda}$$

$$\xrightarrow{\sum \mathrm{pr}^{(j)}_{a_\bullet,!}} H_{\mathrm{et}}^{d(a_\bullet)}\big(\overline{\mathrm{Sh}}_{a_\bullet},\mathcal{L}_\xi\big(\tfrac12 d(a_\bullet)\big)\big)_\pi^{\mathrm{Frob}_{p^f}=\lambda}$$

is an isomorphism, where $\mathrm{pr}^{(j)}_{a_\bullet,!}$ is the Gysin map (2.10.1) *and the superscript $\mathrm{Frob}_{p^f}=\lambda$ means taking the* (*direct sum of*) *generalized Frob_{p^f}-eigenspace with eigenvalues in the Weyl group orbit*

$$\lambda\circ\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})\cdot\chi^{-1}_{\pi_{p,0}}(p^f)(\sqrt{p})^{-f(n-1)\sum_i b_i}.$$

Here, since the semisimple conjugacy classes of $\mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$ is in natural bijection with the orbits of $T(\overline{\mathbb{Q}}_\ell)$ under the Weyl group of GL_n, it makes sense to evaluate a dominant weight of T on $\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})$ to get an orbit under the action of the Weyl group of GL_n; hence the notation $\lambda\circ\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})$.

In particular, when ξ is the trivial representation and the weight λ is a power of the determinant (*so automatically, $\sum_i a_i$ is divisible by n, and $d(a_\bullet)$ is even*), *the cycles given by the images of $Y_1,\dots,Y_{m_\lambda(a_\bullet)}$ parameterized by the* ***discrete*** *Shimura varieties $\mathrm{Sh}_{b_\bullet^{(j)}}$, generate the Tate classes of $H_{\mathrm{et}}^{d(a_\bullet)}\big(\overline{\mathrm{Sh}}_{a_\bullet},\overline{\mathbb{Q}}_\ell\big(\tfrac12 d(a_\bullet)\big)\big)_\pi$ when $\rho_{\pi_\mathfrak{p}}(\mathrm{Frob}_{p^f})$ is "sufficiently generic".*

[9] This isomorphism depends on the choice of the isomorphism $\gamma_{a_\bullet,b_\bullet}$ made earlier.

[10] This assumption is satisfied when π is the finite part of an automorphic cuspidal representation of $G_{a_\bullet}(\mathbb{A})$ which admits a base change to a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_E)\times\mathbb{A}_E^\times$. Indeed, in this case, White [2012, Theorem E] proved that $m_{a_\bullet}(\pi)=m_{b_\bullet^{(j)}}(\pi)=1$.

[11] The π-isotypic component is the same as the π^p-isotypic component according to Lemma 4.17.

Remark 2.13. (1) A key feature of this conjecture is that the codimension of the cycle map $\mathrm{pr}_{a_\bullet}: Y_j \to \mathrm{Sh}_{a_\bullet}$ is the same as the fiber dimension of $\mathrm{pr}_{b_\bullet^{(j)}}: Y_j \to \mathrm{Sh}_{b_\bullet^{(j)}}$.

(2) It seems that the fiber of $\mathrm{pr}_{b_\bullet^{(j)}}: Y_j \to \mathrm{Sh}_{b_\bullet^{(j)}}$ over a generic point $\eta \in \mathrm{Sh}_{b_\bullet^{(j)}}$ is likely to be isomorphic to a certain "iterated Deligne–Lusztig variety," that is, a tower of maps $Y_{j,\eta} = Z_\alpha \to \cdots \to Z_0 = \eta$ such that each $Z_i \to Z_{i-1}$ is a fiber bundle with certain Deligne–Lusztig varieties as fibers.

(3) Xinwen Zhu pointed out to us that since the universal abelian varieties $\mathcal{A}_{a_\bullet}$ and $\mathcal{A}_{b_\bullet}$ are isogenous over each Y_j, the union of the images of $Y_1, \dots, Y_{m_\lambda(a_\bullet)}$ on $\mathrm{Sh}_{a_\bullet}$ is contained in the closure of the Newton strata, where the slope is the same as the μ-ordinary slope of the universal abelian varieties on $\mathrm{Sh}_{b_\bullet^{(j)}}$ (for different j, they have the same μ-ordinary slopes). In fact, one should expect the union of images to be the same as the closure of this Newton stratum.

When λ is central (i.e., a power of the determinant), Conjecture 2.12 says: *irreducible components of the basic locus of the special fiber of a Shimura variety, generically, contribute to all Tate cycles in the cohomology*. Implicitly, this means that the dimension of the basic locus is half of the dimension of the Shimura variety if and only if the Galois representations of the Shimura variety has generically nontrivial Tate classes. Here two appearances of "generic" both mean that we only consider those π-isotypic components where the Satake parameter for π_p is sufficiently generic as in Conjecture 2.12(4). For example, the supersingular locus of Hilbert modular surface at a split prime or the supersingular locus of a Siegel modular variety (over $\mathbb{Q}$) is not half the dimension. This is related to the fact that the π-isotypic component of the cohomology of the Shimura varieties are not expected to have Tate classes, at least when the Satake parameter of π_p is sufficiently general.[12]

(4) These varieties Y_j may be viewed as Hecke correspondences at p *between the special fibers of two different Shimura varieties* $\mathrm{Sh}_{a_\bullet}$ and $\mathrm{Sh}_{b_\bullet^{(j)}}$. These correspondences certainly cannot be lifted to characteristic zero. We hope that the conjecture will bring interests into the study of such Hecke correspondences.

Remark 2.14. (1) The assumption on the decomposition of the place p in $E/\mathbb{Q}$ and working with unitary Shimura varieties is to simplify our presentation and to get to a situation where most terms can be defined. We certainly expect the validity of analogous conjectures for the special fibers of Shimura varieties of PEL-type or more generally of abelian type (using the integral model of M. Kisin [2010]). This would be a more precise version of the Tate conjecture in the context of special fibers of Shimura varieties: if Sh_G and $\mathrm{Sh}_{G'}$ are the

[12] The Siegel varieties are Shimura varieties associated to $\mathrm{GSp}_{2g}(\mathbb{Q})$. The Langlands dual group is isogenous to $\mathrm{Spin}(2g+1)$ and the associated representation r_μ is the spin representation, which is minuscule and hence does not contain trivial weight subspace.

special fibers of two unitary Shimura varieties associated to the groups G and G' such that $G(\mathbb{A}_f) \simeq G'(\mathbb{A}_f)$, then, generically, the cycles on the product $\mathrm{Sh}_G \times \mathrm{Sh}_{G'}$ predicted by the Tate conjectures are likely to be constructed by understanding the "isogenies" between the corresponding universal abelian varieties, and are closely related to the Newton stratifications of Sh_G and $\mathrm{Sh}_{G'}$. In the case of Shimura varieties of abelian type, we expect some technical difficulties in reinterpreting the meaning of isogenies of abelian varieties in terms of certain "G-crystals".

For example, consider a real quadratic field $F/\mathbb{Q}$ in which a prime p is inert. Let Sh_G denote the special fiber of the Hilbert–Siegel modular variety for $G := \mathrm{Res}_{F/\mathbb{Q}} \mathrm{GSp}_{2g}$, with hyperspecial level structure at p. Then by Langlands's prediction of the cohomology of Sh_G, we should look at the representation $r_{\mathrm{spin}}^{\otimes 2}$ of the "essential part" Spin_{2g+1} of the Langlands dual group, where r_{spin} is the 2^g-dimensional spin representation.[13] The central weight space of $r_{\mathrm{spin}}^{\otimes 2}$ has dimension 2^g. So we expect that the supersingular locus of Sh_G is the union of 2^g collection of varieties parameterized by the discrete Shimura variety $\mathrm{Sh}_{G'}$ where G' is the inner form of G which is split at all finite places and is compact modulo center at both archimedean places. Unfortunately, the moduli problem that describes G' uses a different division algebra from that describing G. We do not know how to interpret the meaning of isogenies of universal abelian varieties in this case, and the method of our paper does not apply directly to this case.

(2) Xinwen Zhu pointed out to us that even if p is ramified, we should expect Conjecture 2.12 continue to hold for (the special fiber of) the "splitting models" of Pappas and Rapoport [2005]. Some evidences of this have already appeared in the case of Hilbert modular varieties; see [Rapoport et al. 2014; Reduzzi and Xiao 2017].

(3) In our setup, we took advantage of many coincidences that ensures that for example the Shimura variety is compact and there is no endoscopy. It would be certainly an interesting future question to study the case involving Eisenstein series, as well as the case when the representations come from endoscopy transfers.

(4) As explained in Remark 2.13(3), the images of Y_j are expected to form the closure of a certain Newton polygon where the slopes are related to λ. Conjecture 2.12(1)–(3) may have a degenerate situation: when $\sum_i a_i$ is not divisible by n, the representation $V_{a_\bullet}$ does not contain a weight corresponding to a power of the determinant (which corresponds to the basic locus). So our

[13] As pointed out above, we have to work with the Hilbert–Siegel setup as opposed to the usual Siegel setup because r_{spin} is a minuscule representation.

conjecture does not describe the basic locus of $\mathrm{Sh}_{a_\bullet}$, and it is indeed not of half dimension of $\mathrm{Sh}_{a_\bullet}$.

Yet, this basic locus may still have a good description as the union of some fiber bundles over the special fibers of some other Shimura varieties for reductive groups which are *not* quasisplit at p. For example, the supersingular locus of modular curve is related to the Shimura variety associated to the definite quaternion algebra which is ramified at p, by a theorem of Serre and Deuring [Serre 1996]. More such examples are given in [Tian and Xiao 2016] and [Vollaard and Wedhorn 2011].

2.15. ***Known cases of Conjecture 2.12.*** Conjecture 2.12 is largely inspired by the work of Tian and Xiao [2014; 2016], where we proved the analogous conjecture for the special fibers of the Hilbert modular varieties assuming that p is inert in the totally real field.

Another strong evidence of Conjecture 2.12 is the work of Vollaard and Wedhorn [2011], where they considered certain stratification of the supersingular locus of the Shimura variety for $GU(1, n-1)$ with $s \in \mathbb{N}$ at an *inert* prime p. What concerns us is the case when $n-1$ is even. In this case, it is hidden in the writing of their Section 6 that one gets a correspondence (in the notation of *loc. cit.*)

$$\begin{tikzcd}[column sep=small] & I(\mathbb{Q}) \backslash N_n \times^{C_p^{(n)}} J(\mathbb{Q}_p) \times \mathbb{G}(\mathbb{A}_f^{(p)})/C^p \arrow[dl] \arrow[dr] & \\ I(\mathbb{Q}) \backslash I(\mathbb{A}_f)/C^p C_p^{(n)} & & \mathcal{M}^{ss}_{C^p} \subset \mathcal{M}_{C^p}. \end{tikzcd} \tag{2.15.1}$$

Note that $I(\mathbb{A}_f) \simeq \mathbb{G}(\mathbb{A}_f)$. Here N_n is a certain Deligne–Lusztig variety. In [Vollaard and Wedhorn 2011], the parameterizing space, namely the first term in (2.15.1), is interpreted very differently, in terms of Bruhat–Tits building. The method of this paper should be applicable to their situation to verify the analogous Conjecture 2.12. In fact, in their case, there will be only one collection of cycles as given by (2.15.1), but the computation of the intersection matrix (only essentially one entry in this case) of them requires some nontrivial Schubert calculus similar to Section 5.

When $n-1$ is odd, the result of [Vollaard and Wedhorn 2011] is related to the degenerate version of the Conjecture 2.12 in the sense of Remark 2.14(4).

The aim of the rest of the paper is to provide evidence for Conjecture 2.12 for some large rank groups. In particular, we will construct cycles in the case of the unitary group $G(U(r,s) \times U(s,r))$ with $s, r \in \mathbb{N}$ (Section 7). While we expect these cycles to verify Conjecture 2.12, we do not know how to compute the "intersection matrix" in general. Nonetheless, when $r = 1$, we are able to make the computation and prove Conjecture 2.12 (with trivial coefficients for the sake of a simple presentation) in this case; see Section 4–6. We point out that our method should be applicable to many other examples, and even in general reduce Conjecture 2.12 to a question of a combinatorial nature. This combinatorics problem is the heart of the question. In the Hilbert case [Tian and Xiao 2014], we model the combinatorics question by the so-called periodic semimeander (for GL_2). The generalization of the usual (as opposed to periodic) semimeander to other groups has been introduced; see [Fontaine et al. 2013] for the corresponding references. The straightforward generalization to the periodic case does seem to agree with some of our computations with small groups. Nonetheless, the corresponding Gram determinant formula seems to be extremely difficult. Even in the nonperiodic case, we only know it for a special case; see [Di Francesco 1997].

We also mention that in a very recent work [Xiao and Zhu 2017] of Zhu and the last author, we relate Conjecture 2.12 with the geometric Satake theory of Zhu [2017] in mixed characteristic, and we proved many new cases of Conjecture 2.12.

3. Preliminaries on Dieudonné modules and deformation theory

We first introduce the basic tools that we will use in this paper.

3.1. ***Notation.*** Recall that we have an isomorphism

$$\mathcal{O}_D\otimes_{\mathbb{Z}}\mathbb{Z}_{p^f}\cong\bigoplus_{i=1}^{f}(\mathcal{O}_D\otimes_{\mathcal{O}_E,q_i}\mathbb{Z}_{p^f}\oplus\mathcal{O}_D\otimes_{\mathcal{O}_E,\bar{q}_i}\mathbb{Z}_{p^f})\simeq\bigoplus_{i=1}^{f}(\mathrm{M}_n(\mathbb{Z}_{p^f})\oplus\mathrm{M}_n(\mathbb{Z}_{p^f})).$$

Let S be a locally noetherian $\mathbb{Z}_{p^f}$-scheme. An $\mathcal{O}_D\otimes_{\mathbb{Z}}\mathcal{O}_S$-module M admits a canonical decomposition

$$M=\bigoplus_{i=1}^{f}(M_{q_i}\oplus M_{\bar{q}_i}),$$

where M_{q_i} (resp. $M_{\bar{q}_i}$) is the direct summand of M on which $\mathcal{O}_E$ acts via q_i (resp. via $\bar{q}_i$). Then each M_{q_i} has a natural action by $\mathrm{M}_n(\mathcal{O}_S)$. Let $\mathfrak{e}$ denote the element of $\mathrm{M}_n(\mathcal{O}_S)$ whose $(1,1)$-entry is 1 and whose other entries are 0. We put $M_i^\circ:=\mathfrak{e}M_{q_i}$, and call it the *reduced part* of M_{q_i}.

Let A be an fn^2-dimensional abelian variety over an $\mathbb{F}_{p^f}$-scheme S, equipped with an $\mathcal{O}_D$-action. The de Rham homology $H_1^{\mathrm{dR}}(A/S)$ has a Hodge filtration

$$0\to\omega_{A^\vee/S}\to H_1^{\mathrm{dR}}(A/S)\to\mathrm{Lie}_{A/S}\to 0,$$

compatible with the natural action of $\mathcal{O}_D\otimes_{\mathbb{Z}}\mathcal{O}_S$ on $H_1^{\mathrm{dR}}(A/S)$. We call $H_1^{\mathrm{dR}}(A/S)_i^\circ$ (resp. $\omega^\circ_{A^\vee/S,i}$, $\mathrm{Lie}^\circ_{A/S,i}$) the *reduced de Rham homology* of A/S (resp. the *reduced invariant* 1*-forms* of $A^\vee/S$, the *reduced Lie algebra* of A/S) at q_i. In particular, the former is a locally free $\mathcal{O}_S$-module of rank n and the latter is a subbundle[14] of the former; when $A \to S$ satisfies the moduli problem in Section 2.3, $\omega^\circ_{A^\vee/S,i}$ is locally free of rank a_i.

The Frobenius morphism $A \to A^{(p)}$ induces a natural homomorphism

$$V : H_1^{\mathrm{dR}}(A/S)^\circ_i \to H_1^{\mathrm{dR}}(A/S)^{\circ,(p)}_{i-1},$$

where the index i is considered as an element of $\mathbb{Z}/f\mathbb{Z}$, and the superscript "(p)" means the pullback via the absolute Frobenius of S. The image of V is exactly $\omega^{\circ,(p)}_{A^\vee/S,i-1}$. Similarly, the Verschiebung morphism $A^{(p)} \to A$ induces a natural homomorphism[15]

$$F : H_1^{\mathrm{dR}}(A/S)^{\circ,(p)}_{i-1} \to H_1^{\mathrm{dR}}(A/S)^\circ_i.$$

We have $\mathrm{Ker}(F) = \mathrm{Im}(V)$ and $\mathrm{Ker}(V) = \mathrm{Im}(F)$.

When $S = \mathrm{Spec}(k)$ with k a perfect field containing $\mathbb{F}_{p^f}$, let $W(k)$ denote the ring of Witt vectors in k. Let $\tilde{\mathcal{D}}(A)$ denote the (covariant) Dieudonné module associated to the p-divisible group of A. This is a free $W(k)$-module of rank $2fn^2$ equipped with a Frob-linear action of F and a Frob^{-1}-linear action of V such that $FV = VF = p$. The $\mathcal{O}_D$-action on A induces a natural action of $\mathcal{O}_D$ on $\tilde{\mathcal{D}}(A)$ that commutes with F and V. Moreover, there is a canonical isomorphism $\tilde{\mathcal{D}}(A)/p\tilde{\mathcal{D}}(A) \cong H_1^{\mathrm{dR}}(A/k)$ compatible with all structures on both sides. For each $i \in \mathbb{Z}/f\mathbb{Z}$, we have the reduced part $\tilde{\mathcal{D}}(A)^\circ_i := \mathfrak{e}\tilde{\mathcal{D}}(A)_{q_i}$. The Verschiebung and the Frobenius induce natural maps

$$V : \tilde{\mathcal{D}}(A)^\circ_i \to \tilde{\mathcal{D}}(A)^\circ_{i-1}, \quad F : \tilde{\mathcal{D}}(A)^\circ_i \to \tilde{\mathcal{D}}(A)^\circ_{i+1}.$$

Note that $\tilde{\mathcal{D}}(A)_{q_i} = (\tilde{\mathcal{D}}(A)^\circ_i)^{\oplus n}$, and $\bigoplus_{i \in \mathbb{Z}/f\mathbb{Z}} \tilde{\mathcal{D}}(A)_{q_i}$ is the covariant Dieudonné module of the p-divisible group $A[\mathfrak{p}^\infty]$.

For any fn^2-dimensional abelian variety A' over k equipped with an $\mathcal{O}_D$-action, an $\mathcal{O}_D$-equivariant isogeny $A' \to A$ induces a morphism $\tilde{\mathcal{D}}(A')^\circ_i \to \tilde{\mathcal{D}}(A)^\circ_i$ compatible with the actions of F and V. Conversely, we have the following.

Proposition 3.2. *Let A be an abelian variety of dimension fn^2 over prefect field k which contains $\mathbb{F}_{p^f}$, equipped with an $\mathcal{O}_D$-action and an $\mathcal{O}_D$-compatible prime-to-p polarization λ. Suppose given an integer $m \geq 1$ and a $W(k)$-submodule $\tilde{\mathcal{E}}_i \subseteq \tilde{\mathcal{D}}(A)^\circ_i$ for each $i \in \mathbb{Z}/f\mathbb{Z}$ such that*

$$p^m\tilde{\mathcal{D}}(A)^\circ_i \subseteq \tilde{\mathcal{E}}_i, \quad F(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{i+1}, \quad \textit{and} \quad V(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{i-1}. \tag{3.2.1}$$

[14] Here and after, by a subbundle of a locally free coherent sheaf, we mean a locally free coherent sheaf that is Zariski locally a direct factor.

[15] The notation F for Frobenius was also used to denote the real quadratic field. But we think the chance for confusion is minimal.

Then there exists a unique abelian variety A' over k (depending on m) equipped with an $\mathcal{O}_D$-action, a prime-to-p polarization λ', and an $\mathcal{O}_D$-equivariant p-isogeny $\phi : A' \to A$ such that the natural inclusion $\tilde{\mathcal{E}}_i \subseteq \tilde{\mathcal{D}}(A)^\circ_i$ is naturally identified with the map $\phi_{*,i} : \tilde{\mathcal{D}}(A')^\circ_i \to \tilde{\mathcal{D}}(A)^\circ_i$ induced by ϕ and such that $\phi^\vee \circ \lambda \circ \phi = p^m \lambda'$. Moreover, we have

(1) *If* $\dim \omega^\circ_{A^\vee/k,i} = a_i$ *and* $\mathrm{length}_{W(k)}(\tilde{\mathcal{D}}(A)^\circ_i/\tilde{\mathcal{E}}_i) = \ell_i$ *for $i \in \mathbb{Z}/f\mathbb{Z}$, then*

$$\dim \omega^\circ_{A'^\vee/k,i} = a_i + \ell_i - \ell_{i+1}. \tag{3.2.2}$$

(2) *If A is equipped with a prime-to-p level structure η in the sense of Section 2.3(1), then there exists a unique prime-to-p level structure η' on A' such that $\eta = \phi \circ \eta'$.*

Proof. By Dieudonné theory, the Dieudonné submodule

$$\bigoplus_{i \in \mathbb{Z}/f\mathbb{Z}} (\tilde{\mathcal{E}}_i/p^m \tilde{\mathcal{D}}(A)^\circ_i)^{\oplus n} \subseteq \bigoplus_{i \in \mathbb{Z}/f\mathbb{Z}} (\tilde{\mathcal{D}}(A)^\circ_i/p^m \tilde{\mathcal{D}}(A)^\circ_i)^{\oplus n}$$

corresponds to a closed subgroup scheme $H_{\mathfrak{p}} \subseteq A[\mathfrak{p}^m]$. The prime-to-$p$ polarization λ induces a perfect pairing

$$\langle -, - \rangle_\lambda : A[\mathfrak{p}^m] \times A[\bar{\mathfrak{p}}^m] \to \mu_{p^m}.$$

Let $H_{\bar{\mathfrak{p}}} = H_{\mathfrak{p}}^\perp \subseteq A[\bar{\mathfrak{p}}^m]$ denote the orthogonal complement of $H_{\mathfrak{p}}$. Put $H_p = H_{\mathfrak{p}} \oplus H_{\bar{\mathfrak{p}}}$. Let $\psi : A \to A'$ be the canonical quotient with kernel H_p, and $\phi : A' \to A$ be the quotient with kernel $\psi(A[\mathfrak{p}^m])$ so that $\psi \circ \phi = p^m \,\mathrm{id}_{A'}$ and $\phi \circ \psi = p^m \,\mathrm{id}_A$. By construction, $H_p \subseteq A[p^m]$ is a maximal totally isotropic subgroup. By [Mumford 2008, §23, Theorem 2], there is a prime-to-p polarization λ' on A' such that $p^m \lambda = \psi^\vee \circ \lambda' \circ \psi$. It follows also that $p^m \lambda' = \phi^\vee \circ \lambda \circ \phi$. The fact that $\phi_{*,i} : \tilde{\mathcal{D}}(A')^\circ_i \to \tilde{\mathcal{D}}(A)^\circ_i$ is identified with the natural inclusion $\tilde{\mathcal{E}}_i \subseteq \tilde{\mathcal{D}}(A)^\circ_i$ follows from the construction. The existence and uniqueness of the tame level structure is clear. The dimension of the differential forms can be computed as follows:

$$\begin{aligned}
\dim_k \omega^\circ_{A'^\vee/k,i} &= \dim_k \frac{V(\tilde{\mathcal{D}}(A')^\circ_{i+1})}{p\tilde{\mathcal{D}}(A')^\circ_i} = \dim_k \frac{V(\tilde{\mathcal{E}}_{i+1})}{p\tilde{\mathcal{E}}_i} \\
&= \dim_k \frac{V(\tilde{\mathcal{D}}(A)^\circ_{i+1})}{p\tilde{\mathcal{D}}(A)^\circ_i} - \mathrm{length}_{W(k)} \frac{V(\tilde{\mathcal{D}}(A)^\circ_{i+1})}{V(\tilde{\mathcal{E}}_{i+1})} + \mathrm{length}_{W(k)} \frac{p\tilde{\mathcal{D}}(A)^\circ_i}{p\tilde{\mathcal{E}}_i} \\
&= a_i - \ell_{i+1} + \ell_i.
\end{aligned}$$

□

3.3. ***Deformation theory.*** We shall frequently use Grothendieck–Messing deformation theory to compare the tangent spaces of moduli spaces. We make this explicit in our setup.

Let $\hat{R}$ be a noetherian $\mathbb{F}_{p^f}$-algebra and $\hat{I} \subset \hat{R}$ an ideal such that $\hat{I}^2 = 0$. Put $R = \hat{R}/\hat{I}$. Let $\mathscr{C}_{\hat{R}}$ denote the category of tuples $(\hat{A}, \hat{\lambda}, \hat{\eta})$, where $\hat{A}$ is an fn^2-dimensional abelian variety over $\hat{R}$ equipped with an $\mathcal{O}_D$-action, $\hat{\lambda}$ is a polarization on $\hat{A}$ such that the Rosati involution induces the $*$-involution on $\mathcal{O}_D$, and $\hat{\eta}$ is a level structure as in Section 2.3(3). We define $\mathscr{C}_R$ in the same way. For an object (A, λ, η) in the category $\mathscr{C}_R$, let $H_1^{\mathrm{cris}}(A/\hat{R})$ be the evaluation of the first relative crystalline homology (i.e., dual crystal of the first crystalline cohomology) of A/R at the divided power thickening $\hat{R} \to R$, and $H_1^{\mathrm{cris}}(A/\hat{R})_i^\circ := \mathfrak{e}H_1^{\mathrm{cris}}(A/\hat{R})_{q_i}$ be the i-th reduced part. We denote by $\mathfrak{Def}(R, \hat{R})$ the category of tuples $(A, \lambda, \eta, (\hat{\omega}_i^\circ)_{i=1,\dots,f})$, where (A, λ, η) is an object in $\mathscr{C}_R$, and $\hat{\omega}_i^\circ \subseteq H_1^{\mathrm{cris}}(A/\hat{R})_i^\circ$ for each $i \in \mathbb{Z}/f\mathbb{Z}$ is a subbundle that lifts $\omega^\circ_{A^\vee/R,i} \subseteq H_1^{\mathrm{dR}}(A/R)_i^\circ$. The following is a combination of Serre–Tate and Grothendieck–Messing deformation theory.

Theorem 3.4 (Serre–Tate, Grothendieck–Messing). *The functor*

$$(\hat{A}, \hat{\lambda}, \hat{\eta}) \mapsto (\hat{A} \otimes_{\hat{R}} R, \lambda, \eta, \omega^\circ_{\hat{A}^\vee/\hat{R},i}),$$

where λ and η are the natural induced polarization and level structure on $\hat{A} \otimes_{\hat{R}} R$, is an equivalence of categories between $\mathscr{C}_{\hat{R}}$ and $\mathfrak{Def}(R, \hat{R})$.

Proof. The main theorem of the crystalline deformation theory (cf., [Grothendieck 1974, pp. 116–118], [Mazur and Messing 1974, Chapter II §1]) says that the category $\mathscr{C}_{\hat{R}}$ is equivalent to the category of objects (A, λ, η) in $\mathscr{C}_R$ together with a lift of $\omega_{A^\vee/R} \subseteq H_1^{\mathrm{cris}}(A/R)$ to a subbundle $\hat{\omega}$ of $H_1^{\mathrm{cris}}(A/\hat{R})$, such that $\hat{\omega}$ is stable under the induced $\mathcal{O}_D$-action and is isotropic for the pairing on $H_1^{\mathrm{cris}}(A/\hat{R})$ induced by the polarization λ. But the additional information $\hat{\omega}$ is clearly equivalent to the subbundles $\hat{\omega}_i^\circ \subseteq H_1^{\mathrm{cris}}(A/\hat{R})_i^\circ$ lifting $\omega^\circ_{A^\vee/R,i}$. □

Corollary 3.5. *If $\mathcal{A}_{a_\bullet}$ denotes the universal abelian variety over $\mathrm{Sh}_{a_\bullet}$, then the tangent space $T_{\mathrm{Sh}_{a_\bullet}}$ of $\mathrm{Sh}_{a_\bullet}$ is*

$$\bigoplus_{i=1}^{f} \mathrm{Lie}^\circ_{\mathcal{A}^\vee_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i} \otimes \mathrm{Lie}^\circ_{\mathcal{A}_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i}.$$

Proof. Even though this is a well-known statement often referred to as the Kodaira–Spencer isomorphism (e.g., [Lan 2013, Proposition 2.3.4.2]), we include a short proof, as the proof serves as a toy model of many arguments later. Let $\hat{R}$ be a noetherian $\mathbb{F}_{p^f}$-algebra and $\hat{I} \subset \hat{R}$ an ideal such that $\hat{I}^2 = 0$; put $R = \hat{R}/\hat{I}$. By Theorem 3.4, to lift an R-point (A, λ, η) of $\mathrm{Sh}_{a_\bullet}$ to an $\hat{R}$-point, it suffices to lift, for $i = 1, \dots, f$, the differentials $\omega^\circ_{A^\vee,i} \subseteq H_1^{\mathrm{cris}}(A/R)_i^\circ$ to a subbundle $\hat{\omega}_i \subseteq H_1^{\mathrm{cris}}(A/\hat{R})_i^\circ$. Such lifts form a torsor for the group

$$\mathrm{Hom}_R(\omega^\circ_{A^\vee/R,i}, \mathrm{Lie}^\circ_{A/R,i}) \otimes_R \hat{I}.$$

It follows from this

$$T_{\mathrm{Sh}_{a_\bullet}} \cong \bigoplus_{i=1}^{f} \mathcal{H}om(\omega^\circ_{\mathcal{A}^\vee_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i}, \mathrm{Lie}^\circ_{\mathcal{A}_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i}) \cong \bigoplus_{i=1}^{f} \mathrm{Lie}^\circ_{\mathcal{A}^\vee_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i} \otimes \mathrm{Lie}^\circ_{\mathcal{A}_{a_\bullet}/\mathrm{Sh}_{a_\bullet},i}.$$

Note that this proof also shows that $\mathrm{Sh}_{a_\bullet}$ is smooth. □

3.6. ***Notation in the real quadratic case.*** For the rest of the paper, we assume $f = 2$ so that F is a real quadratic field in which p is inert. For nonnegative integers $r \leq s$ such that $n = r + s$, we denote by $G_{r,s}$ the algebraic group previously denoted by $G_{a_\bullet}$ with $a_1 = r$ and $a_2 = s$; in particular, $G_{r,s}(\mathbb{R}) = G(U(r, s) \times U(s, r))$. If r', s' is another pair of nonnegative integers such that $n = r' + s'$ and $r' \leq s'$, Lemma 2.9 gives an isomorphism $G_{r,s}(\mathbb{A}^\infty) \simeq G_{r',s'}(\mathbb{A}^\infty)$.

Let $\mathcal{S}h_{r,s}$ be the Shimura variety over $\mathbb{Z}_{p^2}$ attached to $G_{r,s}$ defined in Section 2.3 of some fixed sufficiently small prime-to-p level $K^p \subseteq G_{r,s}(\mathbb{A}^{\infty,p})$. Let $\mathrm{Sh}_{r,s}$ denote its special fiber over $\mathbb{F}_{p^2}$. Let $\mathcal{A} = \mathcal{A}_{r,s}$ denote the universal abelian variety over $\mathrm{Sh}_{r,s}$. It is a $2n^2$-dimensional abelian variety, equipped with an action of $\mathcal{O}_D$ and a prime-to-p polarization $\lambda_{\mathcal{A}}$. Moreover, $\omega^\circ_{\mathcal{A}^\vee/\mathrm{Sh}_{r,s},1}$ (resp. $\omega^\circ_{\mathcal{A}^\vee/\mathrm{Sh}_{r,s},2}$) is a locally free module over $\mathrm{Sh}_{r,s}$ of rank r (resp. rank s).

Remark 3.7. When $r = 0$ and $s = n$, the universal abelian variety $\mathcal{A} = \mathcal{A}_{0,n}$ over $\mathrm{Sh}_{0,n}$ is supersingular. Indeed, for each $\bar{\mathbb{F}}_p$-point z of $\mathrm{Sh}_{0,n}$, the Kottwitz condition implies that the Frobenius induces *isomorphisms*

$$\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \xrightarrow{F} \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2 \xrightarrow{F} p\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1.$$

In particular, $(1/p)F^2$ induces a σ^2-linear automorphism of $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1$. By Hilbert's Theorem 90, there exists a $\mathbb{Z}_{p^2}$-lattice $\mathbb{L}$ of $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1$ that is invariant under the action of $(1/p)F^2$; in other words, F^2 acts by multiplication by p for a basis chosen from this lattice. It follows that all slopes of the Frobenius on $\tilde{\mathcal{D}}(\mathcal{A}_z)$ are $\frac{1}{2}$, and hence $\mathcal{A}_z$ is supersingular.

4. The case of $G(U(1, n-1) \times U(n-1, 1))$

We will verify Conjecture 2.12 for $\mathrm{Sh}_{1,n-1}$, namely the existence of some cycles Y_j having morphisms to both $\mathrm{Sh}_{0,n}$ and $\mathrm{Sh}_{1,n-1}$ and generating Tate classes of $\mathrm{Sh}_{1,n-1}$ under a certain genericity hypothesis on the Satake parameters. We always fix an isomorphism $G_{1,n-1}(\mathbb{A}^\infty) \simeq G_{0,n}(\mathbb{A}^\infty)$, and write $G(\mathbb{A}^\infty)$ for either group.

Notation 4.1. For a smooth variety X over $\mathbb{F}_{p^2}$, we denote by T_X the tangent bundle of X, and for a locally free $\mathcal{O}_X$-module M, we put $M^* = \mathcal{H}om_{\mathcal{O}_X}(M, \mathcal{O}_X)$.

4.2. ***Cycles on*** $\mathbf{Sh}_{1,n-1}$**.** For each integer j with $1 \leq j \leq n$, we first define the variety Y_j we briefly mentioned in the introduction. Let Y_j be the moduli space over $\mathbb{F}_{p^2}$ that associates to each locally noetherian $\mathbb{F}_{p^2}$-scheme S, the set of isomorphism classes of tuples $(A, \lambda, \eta, B, \lambda', \eta', \phi)$, where

- (A, λ, η) is an S-point of $\mathrm{Sh}_{1,n-1}$,
- (B, λ', η') is an S-point of $\mathrm{Sh}_{0,n}$ and
- $\phi : B \to A$ is an $\mathcal{O}_D$-equivariant isogeny whose kernel is contained in $B[p]$,

such that

- $p\lambda' = \phi^\vee \circ \lambda \circ \phi$,
- $\phi \circ \eta' = \eta$ and
- the cokernels of the maps

$$\phi_{*,1} : H_1^{\mathrm{dR}}(B/S)_1^\circ \to H_1^{\mathrm{dR}}(A/S)_1^\circ \quad \text{and} \quad \phi_{*,2} : H_1^{\mathrm{dR}}(B/S)_2^\circ \to H_1^{\mathrm{dR}}(A/S)_2^\circ$$

are locally free $\mathcal{O}_S$-modules of rank $j-1$ and j, respectively.

There is a unique isogeny $\psi : A \to B$ such that $\psi \circ \phi = p \cdot \mathrm{id}_B$ and $\phi \circ \psi = p \cdot \mathrm{id}_A$. We have

$$\mathrm{Ker}(\phi_{*,i}) = \mathrm{Im}(\psi_{*,i}) \quad \text{and} \quad \mathrm{Ker}(\phi_{*,i}) = \mathrm{Im}(\psi_{*,i}),$$

where $\psi_{*,i}$ for $i = 1, 2$ is the induced homomorphism on the reduced de Rham homology in the evident sense. This moduli space Y_j is represented by a scheme of finite type over $\mathbb{F}_{p^2}$. We have a natural diagram of morphisms:

$$\begin{tikzcd} & Y_j \arrow[dl, "\mathrm{pr}_j"'] \arrow[dr, "\mathrm{pr}'_j"] & \\ \mathrm{Sh}_{1,n-1} & & \mathrm{Sh}_{0,n}, \end{tikzcd} \tag{4.2.1}$$

where pr_j and pr'_j send a tuple $(A, \lambda, \eta, B, \lambda', \eta', \phi)$ to (A, λ, η) and to (B, λ', η'), respectively. Letting K^p vary, we see easily that both pr_j and pr'_j are equivariant under prime-to-p Hecke actions given by the double cosets $K^p \backslash G(\mathbb{A}^{\infty,p})/K^p$.

4.3. ***Some auxiliary moduli spaces.*** The moduli problem for Y_j is slightly complicated. We will introduce a more explicit moduli space Y'_j below and then show they are isomorphic.

Consider the functor $\underline{Y}'_j$ which associates to each locally noetherian $\mathbb{F}_{p^2}$-scheme S the set of isomorphism classes of tuples $(B, \lambda', \eta', H_1, H_2)$, where

- (B, λ', η') is an S-valued point of $\mathrm{Sh}_{0,n}$;
- $H_1 \subset H_1^{\mathrm{dR}}(B/S)_1^\circ$ and $H_2 \subset H_1^{\mathrm{dR}}(B/S)_2^\circ$ are $\mathcal{O}_S$-subbundles of rank j and $j-1$ respectively such that

$$V^{-1}(H_2^{(p)}) \subseteq H_1, \quad H_2 \subseteq F(H_1^{(p)}). \tag{4.3.1}$$

Here,

$$F : H_1^{\mathrm{dR}}(B/S)_1^{\circ,(p)} \xrightarrow{\sim} H_1^{\mathrm{dR}}(B/S)_2^\circ \quad \text{and} \quad V : H_1^{\mathrm{dR}}(B/S)_1^\circ \xrightarrow{\sim} H_1^{\mathrm{dR}}(B/S)_2^{\circ,(p)}$$

are respectively the Frobenius and Verschiebung homomorphisms, which are actually isomorphisms because of the signature condition on $\mathrm{Sh}_{0,n}$.

It follows from the moduli problem that the quotients $H_1/V^{-1}(H_2^{(p)})$, $F(H_1^{(p)})/H_2$ are both locally free $\mathcal{O}_S$-modules of rank one.

There is a natural projection $\pi'_j : \underline{Y}'_j \to \mathrm{Sh}_{0,n}$ given by $(B, \lambda', \eta', H_1, H_2) \mapsto (B, \lambda', \eta')$.

Proposition 4.4. *The functor $\underline{Y}'_j$ is representable by a scheme Y'_j smooth and projective over $\mathrm{Sh}_{0,n}$ of dimension $n-1$. Moreover, if $(\mathcal{B}, \lambda', \eta', \mathcal{H}_1, \mathcal{H}_2)$ denotes the universal object over Y'_j, then the tangent bundle of Y'_j is*

$$T_{Y'_j} \cong \big((\mathcal{H}_1/V^{-1}(\mathcal{H}_2^{(p)})\big)^* \otimes \big(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_1^\circ/\mathcal{H}_1\big) \oplus \big(\mathcal{H}_2^* \otimes F(\mathcal{H}_1^{(p)})/\mathcal{H}_2\big).$$

Proof. For each integer m with $0 \le m \le n$ and $i = 1, 2$, let $\boldsymbol{Gr}(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_i^\circ, m)$ be the Grassmannian scheme over $\mathrm{Sh}_{0,n}$ that parametrizes subbundles of the universal de Rham homology $H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_i^\circ$ of rank m. Then $\underline{Y}'_j$ is a closed subfunctor of the product of the Grassmannian schemes

$$\boldsymbol{Gr}(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_1^\circ, j) \times \boldsymbol{Gr}(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_2^\circ, j-1).$$

The representability of $\underline{Y}'_j$ follows. Moreover, Y'_j is projective.

We show now that the structural map $\pi'_j : Y'_j \to \mathrm{Sh}_{0,n}$ is smooth of relative dimension $n-1$. Let $S_0 \hookrightarrow S$ be an immersion of locally noetherian $\mathbb{F}_{p^2}$-schemes with ideal sheaf I satisfying $I^2 = 0$. Suppose we are given a commutative diagram

$$\begin{tikzcd} S_0 \arrow[r, "g_0"] \arrow[d] & Y'_j \arrow[d, "\pi'_j"] \\ S \arrow[r, "h"] \arrow[ur, dashed, "g"] & \mathrm{Sh}_{0,n} \end{tikzcd}$$

with solid arrows. We have to show that, locally for the Zariski topology on S_0, there is a morphism $g : S \to Y'_j$ making the diagram commute. Let B be the abelian scheme over S given by h, and B_0 be the base change to S_0. The morphism g_0 gives rises to subbundles $\overline{H}_1 \subset H_1^{\mathrm{dR}}(B_0/S_0)_1^\circ$ and $\overline{H}_2 \subset H_1^{\mathrm{dR}}(B_0/S_0)_2^\circ$ with

$$F(\overline{H}_1^{(p)}) \supset \overline{H}_2, \quad V^{-1}(\overline{H}_2^{(p)}) \subset \overline{H}_1.$$

Finding g is equivalent to finding a subbundle $H_i \subset H_1^{\mathrm{dR}}(B/S)_i^\circ$ which lifts each $\overline{H}_i$ for $i = 1, 2$ and satisfies (4.3.1); this is certainly possible when passing to small enough affine open subsets of S_0. Thus $\pi'_j : Y'_j \to \mathrm{Sh}_{0,n}$ is formally smooth, and hence smooth. We note that $F_S^* : \mathcal{O}_S \to \mathcal{O}_S$ factors through $\mathcal{O}_{S_0}$. Hence $V^{-1}(H_2^{(p)})$ and $F(H_1^{(p)})$ actually depend only on $\overline{H}_1, \overline{H}_2$, but not on the lifts H_1 and H_2. Therefore, the possible lifts H_2 form a torsor under the group

$$\mathcal{H}om_{\mathcal{O}_{S_0}}(\overline{H}_2, F(\overline{H}_1^{(p)})/\overline{H}_2) \otimes_{\mathcal{O}_{S_0}} I,$$

and similarly the possible lifts H_1 form a torsor under the group

$$\mathcal{H}om_{\mathcal{O}_{S_0}}(\overline{H}_1/V^{-1}(\overline{H}_2^{(p)}),\, H_1^{\mathrm{dR}}(B_0/S_0)_1^\circ/\overline{H}_1)\otimes_{\mathcal{O}_{S_0}} I.$$

To compute the tangent bundle $T_{Y'_j}$, we take $S=\mathrm{Spec}(\mathcal{O}_{S_0}[\epsilon]/\epsilon^2)$ and $I=\epsilon\mathcal{O}_S$. The morphism $g_0: S_0\to Y'_j$ corresponds to an S_0-valued point of Y'_j, say y_0. Then the possible liftings g form the tangent space $T_{Y'_j}$ at y_0, denote by $T_{Y'_j,y_0}$. The discussion above shows that

$$T_{Y'_j,y_0}\cong\mathcal{H}om_{\mathcal{O}_{S_0}}(\overline{H}_2,F(\overline{H}_1^{(p)})/\overline{H}_2)\oplus\mathcal{H}om_{\mathcal{O}_{S_0}}(\overline{H}_1/V^{-1}(\overline{H}_2^{(p)}),H_1^{\mathrm{dR}}(B_0/S_0)_1^\circ/\overline{H}_1),$$

which is certainly a vector bundle over S_0 of rank $j-1+(n-j)=n-1$. Applying this to the universal case when $g_0: S_0\to Y'_j$ is the identity morphism, the formula of the tangent bundle follows. $\square$

Remark 4.5. Let $(B,\lambda',\eta',H_1,H_2)$ be an S-point of Y'_j.

(a) If $j=n$, H_1 has to be $H_1^{\mathrm{dR}}(B/S)_1^\circ$, and H_2 is a hyperplane of $H_1^{\mathrm{dR}}(B/S)_2^\circ$. Condition (4.3.1) is trivial. In this case, Y'_n is the projective space over $\mathrm{Sh}_{0,n}$ associated to $H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_2^\circ$, where $\mathcal{B}$ is the universal abelian scheme over $\mathrm{Sh}_{0,n}$. So it is geometrically a union of copies of $\mathbb{P}^{n-1}_{\bar{\mathbb{F}}_p}$.

(b) If $j=1$, then H_1 is a line in $H_1^{\mathrm{dR}}(B/S)_1^\circ$ and $H_2=0$. So Y'_1 is the projective space over $\mathrm{Sh}_{0,n}$ associated to $(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_1^\circ)^*$.

(c) If $j=2$, $H_2\subseteq H_1^{\mathrm{dR}}(B/S)_2^\circ$ is a line, and $H_1\subseteq H_1^{\mathrm{dR}}(B/S)_1^\circ$ is a subbundle of rank 2 such that $F(H_1^{(p)})$ contains both H_2 and $F(V^{-1}(H_2^{(p)})^{(p)})$. Therefore, if $H_2\neq F(V^{-1}(H_2^{(p)}))^{(p)}$, H_1 is determined up to Frobenius pullback. If $H_2=F(V^{-1}(H_2^{(p)})^{(p)})$, then H_1 could be any rank 2 subbundle containing $V^{-1}(H_2^{(p)})$.

We fix a geometric point $z=(B,\lambda',\eta')\in\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. It is possible to find good bases for $H_1^{\mathrm{dR}}(B/\bar{\mathbb{F}}_p)_1^\circ$, $H_1^{\mathrm{dR}}(B/\bar{\mathbb{F}}_p)_2^\circ$ such that $F,V: H_1^{\mathrm{dR}}(B/\bar{\mathbb{F}}_p)_1^\circ\to H_1^{\mathrm{dR}}(B/\bar{\mathbb{F}}_p)_2^\circ$ are both given by the identity matrix. With these choices, we may identify the fiber $Y'_{2,z}=\pi_2'^{-1}(z)$ with a closed subvariety of

$$\boldsymbol{Gr}(\bar{\mathbb{F}}_p^n,2)\times\boldsymbol{Gr}(\bar{\mathbb{F}}_p^n,1).$$

Moreover, one may equip $\boldsymbol{Gr}(\bar{\mathbb{F}}_p^n,1)\cong\mathbb{P}^{n-1}_{\bar{\mathbb{F}}_p}$ with an $\mathbb{F}_{p^2}$-rational structure such that $H_2=F(V^{-1}(H_2^{(p)})^{(p)})$ if and only if $[H_2]\in\mathbb{P}^{n-1}_{\bar{\mathbb{F}}_p}$ is an $\mathbb{F}_{p^2}$-rational point. So $Y'_{2,z}$ is isomorphic to a "Frobenius twisted" blow-up of $\mathbb{P}^{n-1}_{\bar{\mathbb{F}}_p}$ at all of its $\mathbb{F}_{p^2}$-rational points. Here, "Frobenius twisted" means that each irreducible component of the exceptional divisor has multiplicity p. For instance, when $n=3$, each $Y_{2,z}$ is isomorphic to the closed subscheme of $\mathbb{P}^2_{\bar{\mathbb{F}}_p}\times\mathbb{P}^2_{\bar{\mathbb{F}}_p}$ defined by

$$a_1b_1^p+a_2b_2^p+a_3b_3^p=0,\quad a_1^pb_1+a_2^pb_2+a_3^pb_3=0,$$

where $(a_1:a_2:a_3)$ and $(b_1:b_2:b_3)$ are the homogeneous coordinates on the two copies of $\mathbb{P}^2$.

Lemma 4.6. *Let $(A, \lambda, \eta, B, \lambda', \eta', \phi)$ be an S-point of Y_j. Then the image of $\phi_{*,1}$ contains both $\omega^\circ_{A^\vee/S,1}$ and $F(H_1^{\mathrm{dR}}(A/S)_2^{\circ,(p)})$, and the image of $\phi_{*,2}$ is contained in both $\omega^\circ_{A^\vee/S,2}$ and $F(H_1^{\mathrm{dR}}(A/S)_1^{\circ,(p)})$.*

Proof. By the functoriality, $\phi_{2,*}$ sends $\omega^\circ_{B^\vee/S,2}$ to $\omega^\circ_{A^\vee/S,2}$. Since $\omega^\circ_{B^\vee/S,2} = H_1^{\mathrm{dR}}(B/S)_2^\circ$ by the Kottwitz determinant condition, it follows that $\mathrm{Im}(\phi_{*,2})$ in contained in $\omega^\circ_{A^\vee/S,2}$. Similar arguments by considering $\psi_{*,1}$ shows that $\omega^\circ_{A^\vee/S,1} \subseteq \mathrm{Ker}(\psi_{*,1}) = \mathrm{Im}(\phi_{*,1})$. The fact that $\mathrm{Im}(\phi_{*,2})$ is contained in $F(H_1^{\mathrm{dR}}(A/S)_1^{\circ,(p)})$ follows from the commutative diagram

$$\begin{tikzcd} H_1^{\mathrm{dR}}(B/S)_1^{\circ,(p)} \arrow[r, "\phi_{*,1}^{(p)}"] \arrow[d, "F"', "\cong"] & H_1^{\mathrm{dR}}(A/S)_1^{\circ,(p)} \arrow[d, "F"] \\ H_1^{\mathrm{dR}}(B/S)_2^\circ \arrow[r, "\phi_{*,2}"] & H_1^{\mathrm{dR}}(A/S)_2^\circ \end{tikzcd} \tag{4.6.1}$$

and the fact that the left vertical arrow is an isomorphism. Similarly, the inclusion $F(H_1^{\mathrm{dR}}(A/S)_2^{\circ,(p)}) \subseteq \mathrm{Im}(\phi_{*,1}) = \mathrm{Ker}(\psi_{*,1})$ can be proved using the functoriality of Verschiebung homomorphisms. □

4.7. ***A morphism from Y_j to Y'_j.*** There is a natural morphism $\alpha : Y_j \to Y'_j$ for $1 \le j \le n$ defined as follows. For a locally noetherian $\mathbb{F}_{p^2}$-scheme S and an S-point $(A, \lambda, \eta, B, \lambda', \eta', \phi)$ of Y_j, we define

$$H_1 := \phi_{*,1}^{-1}(\omega^\circ_{A^\vee/S,1}) \subseteq H_1^{\mathrm{dR}}(B/S)_1^\circ, \quad H_2 := \psi_{*,2}(\omega^\circ_{A^\vee/S,2}) \subseteq H_1^{\mathrm{dR}}(B/S)_2^\circ. \tag{4.7.1}$$

In particular, H_1 and H_2 are $\mathcal{O}_S$-subbundles of rank j and $j-1$, respectively. Also, there is a canonical isomorphism $\omega^\circ_{A^\vee/S,2}/\mathrm{Im}(\phi_{*,2}) \xrightarrow{\sim} H_2$. From the commutative diagram (4.6.1), it is easy to see that $F(H_1^{(p)}) \subseteq \mathrm{Ker}(\phi_{*,2}) = \mathrm{Im}(\psi_{*,2})$, but comparing the rank forces this to be an equality. It follows that $H_2 \subseteq F(H_1^{(p)})$. Similarly, $V^{-1}(H_2^{(p)})$ is identified with $\mathrm{Im}(\psi_{*,1}) = \mathrm{Ker}(\phi_{*,1})$, hence $V^{-1}(H_2^{(p)}) \subseteq H_1$. From these, we deduce two canonical isomorphisms:

$$\begin{aligned} H_1/V^{-1}(H_2^{(p)}) &\xrightarrow{\sim} \omega^\circ_{A^\vee/S,1}, \\ F(H_1^{(p)})/H_2 &\xrightarrow{\sim} H_1^{\mathrm{dR}}(A/S)_2^\circ/\omega^\circ_{A^\vee/S,2} \cong \mathrm{Lie}^\circ_{A/S,2}. \end{aligned} \tag{4.7.2}$$

Therefore, we have a well-defined map $\alpha : Y_j \to Y'_j$ given by

$$\alpha : (A, \lambda, \eta, B, \lambda', \eta', \phi) \mapsto (B, \lambda', \eta', H_1, H_2).$$

Moreover, it is clear from the definition that $\pi'_j \circ \alpha = \mathrm{pr}'_j$.

Proposition 4.8. *The morphism α is an isomorphism.*

Proof. Let k be a perfect field containing $\mathbb{F}_{p^2}$. We first prove that α induces a bijection of points $\alpha : Y_j(k) \xrightarrow{\sim} Y'_j(k)$. It suffices to show that there exists a morphism of sets $\beta : Y'_j(k) \to Y_j(k)$ inverse to α. Let $y = (B, \lambda', \eta', H_1, H_2) \in Y'_j(k)$. We define $\beta(y) = (A, \lambda, \eta, B, \lambda', \eta', \phi)$ as follows. Let $\tilde{\mathcal{E}}_1 \subseteq \tilde{\mathcal{D}}(B)^\circ_1$ and $\tilde{\mathcal{E}}_2 \subseteq \tilde{\mathcal{D}}(B)^\circ_2$ be respectively the inverse images of $V^{-1}(H_2^{(p)}) \subseteq H_1^{\mathrm{dR}}(B/k)^\circ_1$ and $F(H_1^{(p)}) \subseteq H_1^{\mathrm{dR}}(B/k)^\circ_2$ under the natural reduction maps

$$\tilde{\mathcal{D}}(B)^\circ_i \to \tilde{\mathcal{D}}(B)^\circ_i / p\tilde{\mathcal{D}}(B)^\circ_i \cong H_1^{\mathrm{dR}}(B/k)^\circ_i \quad \text{for } i = 1, 2.$$

The condition (4.3.1) ensures that $F(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{3-i}$ and $V(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{3-i}$ for $i = 1, 2$. Applying Proposition 3.2 with $m = 1$, we get a triple (A, λ, η) and an $\mathcal{O}_D$-equivariant isogeny $\psi : A \to B$, where A is an abelian variety over k with an action of $\mathcal{O}_D$, λ is a prime-to-p polarization on A, and η is a prime-to-p level structure on A, such that $\psi^\vee \circ \lambda' \circ \psi = p\lambda$, $p\eta' = \psi \circ \eta$ and such that $\psi_{*,i} : \tilde{\mathcal{D}}(A)^\circ_i \to \tilde{\mathcal{D}}(B)^\circ_i$ is naturally identified with the inclusion $\tilde{\mathcal{E}}_i \hookrightarrow \tilde{\mathcal{D}}(B)^\circ_i$ for $i = 1, 2$. Moreover, the dimension formula (3.2.2) implies that $\omega^\circ_{A^\vee/k,1}$ has dimension 1, and $\omega^\circ_{A^\vee/k,2}$ has dimension $n-1$. Therefore, (A, λ, η) is a point of $\mathrm{Sh}_{1,n-1}$. Finally, we take $\phi : B \to A$ to be the unique isogeny such that $\phi \circ \psi = p \cdot \mathrm{id}_A$ and $\psi \circ \phi = p \cdot \mathrm{id}_B$. Thus we have $\phi \circ \eta' = \eta$. This finishes the construction of $\beta(y)$. It is direct to check that β is the set theoretic inverse to $\alpha : Y_j(k) \to Y'_j(k)$.

We show now that α induces an isomorphism on the tangent spaces at each closed point; as we have already shown that Y'_j is smooth, it will then follow that α is an isomorphism. Let $x = (A, \lambda, \eta, B, \lambda', \eta', \phi) \in Y_j(k)$ be a closed point. Consider the infinitesimal deformation over $k[\epsilon] = k[t]/t^2$. Note that (B, λ', η') has a unique deformation $(\hat{B}, \hat{\lambda}', \hat{\eta}')$ to $k[\epsilon]$, namely the trivial deformation. By the Serre–Tate and Grothendieck–Messing deformation theory (cf., Theorem 3.4), giving a deformation $(\hat{A}, \hat{\lambda}, \hat{\eta})$ of (A, λ, η) to $k[\epsilon]$ is equivalent to giving free $k[\epsilon]$-submodules $\hat{\omega}^\circ_{A^\vee,i} \subseteq H_1^{\mathrm{cris}}(A/k[\epsilon])^\circ_i$ for $i = 1, 2$ which lift $\omega^\circ_{A^\vee/k,i}$. The isogeny ϕ and the polarization λ deform to an isogeny $\hat{\phi} : \hat{B} \to \hat{A}$ and a polarization $\hat{\lambda} : \hat{A}^\vee \to \hat{A}$ (satisfying $p\hat{\lambda}' = \hat{\phi}^\vee \circ \hat{\lambda} \circ \hat{\phi}$), necessarily unique if they exist, if and only if

$$\hat{\omega}^\circ_{A^\vee,2} \supseteq \phi^{\mathrm{cris}}_{*,2}\big(H_1^{\mathrm{cris}}(B/k[\epsilon])^\circ_2\big) \quad \text{and} \quad \big(\phi^{\mathrm{cris}}_{*,1}(H_1^{\mathrm{cris}}(B/k[\epsilon])^\circ_1)\big)^\vee \subseteq (\hat{\omega}^\circ_{A^\vee,1})^\vee,$$

where the second inclusion comes from the consideration at the embedding $\bar{q}_2$ by taking duality using the polarization λ and is equivalent to $\hat{\omega}^\circ_{A^\vee,1} \subseteq \phi^{\mathrm{cris}}_{*,1}\big(H_1^{\mathrm{cris}}(B/k[\epsilon])^\circ_1\big)$.

As discussed before Proposition 4.8, we have $\mathrm{Ker}(\phi_{*,1}) = V^{-1}(H_2^{(p)})$ and $F(H_1^{(p)}) = \mathrm{Ker}(\phi_{*,2}) = \mathrm{Im}(\psi_{*,2})$. Then according to the relation between $\omega^\circ_{A^\vee/k,i}$ and H_1 in (4.7.1), giving such $\hat{\omega}^\circ_{A^\vee,i}$ for $i = 1, 2$ is equivalent to lifting each H_i to a free $k[\epsilon]$-submodule $\hat{H}_i \subseteq H_1^{\mathrm{dR}}(B/k)^\circ_i \otimes_k k[\epsilon] \cong H_1^{\mathrm{cris}}(B/k[\epsilon])^\circ_i$ for $i = 1, 2$ such that $\hat{H}_1 \supseteq V^{-1}(H_2^{(p)}) \otimes_k k[\epsilon]$ and $\hat{H}_2 \subseteq F(H_1^{(p)}) \otimes_k k[\epsilon]$. This is exactly the description of the tangent space of Y'_j at $\alpha(x)$. This concludes the proof. □

In the sequel, we will always identify Y_j with Y'_j and pr'_j with π'_j. Before proceeding, we prove some results on the structure of $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$.

We turn to the study of the Shimura variety $\mathcal{S}h_{0,n}$. The following proposition was suggested by an anonymous referee of this article.

Proposition 4.9. (1) *The Shimura variety $\mathcal{S}h_{0,n}$ is finite and étale over $\mathbb{Z}_{p^2}$. In particular, the reduction map induces a bijection of geometric points*

$$\mathcal{S}h_{0,n}(\overline{\mathbb{Q}}_p) \xrightarrow{\sim} \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p).$$

(2) *Let $\tilde{x}_i = (\tilde{B}_i, \tilde{\lambda}_i, \tilde{\eta}_i) \in \mathcal{S}h_{0,n}(\overline{\mathbb{Q}}_p)$ for $i = 1, 2$ be two geometric points in characteristic* 0, *and $x_i = (B_i, \lambda_i, \eta_i) \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ be their reductions. Then the reduction map on*

$$\mathrm{Hom}_{\mathcal{O}_D}(\tilde{B}_1, \tilde{B}_2) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}_D}(B_1, B_2)$$

is an isomorphism.

Proof. (1) Let $\tilde{z} \in (\tilde{B}, \tilde{\lambda}, \tilde{\eta}) \in \mathcal{S}h_{0,n}(\mathbb{C})$. Put $H = H_1(\tilde{B}(\mathbb{C}), \mathbb{Q})$. It is a left D-module of rank 1 equipped with an alternating D-Hermitian pairing $\langle -, -\rangle_{\tilde{\lambda}}$ induced by the polarization $\tilde{\lambda}$. Let $(V_{0,n} = D, \langle -, -\rangle_{0,n})$ be the left D-module together with its alternating D-Hermitian pairing as in the definition of $\mathcal{S}h_{0,n}$. By results of Kottwitz [1992b, §8], for every place v of $\mathbb{Q}$, the skew-Hermitian $D_{\mathbb{Q}_v}$-modules $H_{\mathbb{Q}_v}$ and $V_{0,n,\mathbb{Q}_v}$ are isomorphic.[16] Then $\mathrm{End}_{\mathcal{O}_D}(\tilde{B}_{\mathbb{C}})_{\mathbb{Q}}$ consists of the elements of $D^{\mathrm{opp}} = \mathrm{End}_D(H)$ that preserves the complex structure on $H_{1,\mathbb{R}} \simeq V_{0,n,\mathbb{R}}$ induced the Deligne homomorphism by $h : \mathbb{C}^\times \to G_{0,n}(\mathbb{R})$. Since $h(i)$ is necessarily central (because $G^1_{0,n}$ is compact), it follows that $\mathrm{End}_{\mathcal{O}_D}(\tilde{B}_{\mathbb{C}})_{\mathbb{Q}} = D^{\mathrm{opp}}$, and

$$D \otimes_E D^{\mathrm{opp}} \simeq \mathrm{M}_{n^2}(E) \subseteq \mathrm{End}(\tilde{B})_{\mathbb{Q}}.$$

For dimension reasons, the inclusion above is an equality, and $\tilde{B}$ is isogenous to the product of n^2-copies of abelian varieties with complex multiplication by E. Therefore, $\tilde{B}$ is defined over a number field and has potentially good reduction everywhere. This implies that $\mathcal{S}h_{0,n}$ is proper over $\mathbb{Z}_{p^2}$.

To see that $\mathcal{S}h_{0,n}$ is finite and étale over $\mathbb{Z}_{p^2}$, it remains to show its étaleness over $\mathbb{Z}_{p^2}$. But this is clear from the description of its relative differential sheaf in Corollary 3.5, which is trivial as $\mathrm{Lie}_{\mathcal{A}^\vee/\mathcal{S}h_{0,n},1} = \mathrm{Lie}_{\mathcal{A}/\mathcal{S}h_{0,n},2} = 0$ by Kottwitz's determinant condition.

(2) In general, the reduction map

$$\mathrm{Hom}_{\mathcal{O}_D}(\tilde{B}_1, \tilde{B}_2) \hookrightarrow \mathrm{Hom}_{\mathcal{O}_D}(B_1, B_2)$$

is injective. It remains to see that every element $f \in \mathrm{Hom}_{\mathcal{O}_D}(B_1, B_2)$ lifts to a homomorphism $\tilde{f} \in \mathrm{Hom}_{\mathcal{O}_D}(\tilde{B}_1, \tilde{B}_2)$. Note that points $\tilde{x}_1, \tilde{x}_2$ can be viewed over

[16] Note that the two skew-Hermitian forms $(H, \langle -, -\rangle)_{\tilde{\lambda}}$ and $(V_{0,n}, \langle -, -\rangle_{0,n})$ are not necessarily isomorphic over $\mathbb{Q}$. However, they differ at most only by a scalar in F, hence define the same similitude unitary group. See [Kottwitz 1992b, p. 400] for details.

$W(\bar{\mathbb{F}}_p)$. As recalled in Section 3, to show that f lifts to a map $\tilde{f}:\tilde{B}_1\to\tilde{B}_2$, it suffices to see that the induced map on crystalline homology

$$f^*: H_1^{\mathrm{cris}}(B_2/W(\bar{\mathbb{F}}_p))\to H_1^{\mathrm{cris}}(B_1/W(\bar{\mathbb{F}}_p))$$

preserves the Hodge filtrations

$$\omega_{\tilde{B}_i^\vee}\subseteq H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))\cong H_1^{\mathrm{cris}}(B_i/W(\bar{\mathbb{F}}_p)).$$

It is clear that f^* preserves the decomposition

$$H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))=H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))_1\oplus H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))_2$$

according to the two embeddings of $\mathcal{O}_E$ into $W(\bar{\mathbb{F}}_p)$. By the Kottwitz's determinant condition for $\mathcal{S}h_{0,n}$, the Hodge filtrations on $H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))$ are trivial, namely,

$$\omega^\circ_{\tilde{B}_i^\vee/W(\bar{\mathbb{F}}_p),1}=0,\quad\text{and}\quad \omega^\circ_{\tilde{B}_i^\vee/W(\bar{\mathbb{F}}_p),2}=H_1^{\mathrm{dR}}(\tilde{B}_i/W(\bar{\mathbb{F}}_p))_2^\circ\quad\text{for } i=1,2.$$

It is clear now f^* preserves this trivial Hodge filtration, since it does so when tensoring with $\bar{\mathbb{F}}_p$. □

Fix a geometric point $z=(B,\lambda,\eta)\in\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. Put $C=\mathrm{End}_{\mathcal{O}_D}(B)_{\mathbb{Q}}$, and denote by $\dagger$ the Rosati involution on C induced by λ. Let I be the algebraic group over $\mathbb{Q}$ such that

$$I(R)=\{x\in C\otimes_{\mathbb{Q}}R\mid xx^\dagger\in R^\times\},\qquad\text{for all }\mathbb{Q}\text{-algebras } R. \tag{4.9.1}$$

Corollary 4.10. *We have an isomorphism of algebraic groups over* $\mathbb{Q}$: $I\simeq G_{0,n}$.

Proof. Let $\tilde{z}=(\tilde{B},\tilde{\lambda},\tilde{\eta})\in\mathcal{S}h_{0,n}(\overline{\mathbb{Q}}_p)$ denote the unique lift of z according to Proposition 4.9 (1). By 4.9 (2), we have a canonical isomorphism

$$\mathrm{End}_{\mathcal{O}_D}(\tilde{B})_{\mathbb{Q}}\xrightarrow{\sim}\mathrm{End}_{\mathcal{O}_D}(B)_{\mathbb{Q}}=C.$$

In the proof of 4.9, we have seen that $C=D^{\mathrm{opp}}$. Moreover, the Rosati involution on C corresponds to the involution $b\mapsto b^{\sharp_{\beta_{0,n}}}=\beta_{0,n}b^*\beta_{0,n}^{-1}$ on D^{opp}, where $\beta_{0,n}$ is the element in the definition of $\langle -,-\rangle_{0,n}$. It follows immediately that $I\simeq G_{0,n}$. □

Let $\mathrm{Isog}(z)\subseteq\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ denote the subset of points $z'=(B',\lambda',\eta')$ such that there exists an $\mathcal{O}_D$-equivariant quasi-isogeny $\phi:B'\to B$ such that $\phi^\vee\circ\lambda\circ\phi=c_0\lambda'$ for some $c_0\in\mathbb{Q}_{>0}$. We denote such a quasi-isogeny by $\phi:z'\to z$ for simplicity.

Corollary 4.11. *There exists a natural bijection of sets*

$$\Theta_z:\mathrm{Isog}(z)\xrightarrow{\sim}G_{0,n}(\mathbb{Q})\backslash G_{0,n}(\mathbb{A}^\infty)/K$$

Proof. First, we give the construction of Θ_z. Put $V^{(p)}(B)=T^{(p)}(B)\otimes_{\hat{\mathbb{Z}}^{(p)}}\mathbb{A}^{\infty,p}$. Then η determines an isomorphism

$$\tilde{\eta}:V_{0,n}^{(p)}\otimes_{\mathbb{Q}}\mathbb{A}^{\infty,p}\xrightarrow{\sim}V^{(p)}(B),$$

modulo right translation by K^p. For any $z'=(B',\lambda',\eta')\in \mathrm{Isog}(z)$ and a choice of $\phi : B'\to B$ as above. The quasi-isogeny ϕ induces an isomorphism $\phi_* : V^{(p)}(B')\xrightarrow{\sim} V^{(p)}(B)$. Then there exists a $g^p\in G_{0,n}(\mathbb{A}^{\infty,p})$, unique up to right multiplication by elements of K^p, such that the K^p-orbit of $\phi_*^{-1}\circ\tilde{\eta}\circ g$ gives η'.

We put

$$\mathbb{L}_z=\tilde{\mathcal{D}}(B)_1^{\circ,F^2=p}=\{v\in\tilde{\mathcal{D}}(B)_1^{\circ} : F^2(v)=pv\}. \tag{4.11.1}$$

Since B is supersingular (See Remark 3.7), this is a free $\mathbb{Z}_{p^2}$-module of rank n, and we have $\tilde{\mathcal{D}}(B)_1^{\circ}=\mathbb{L}_z\otimes_{\mathbb{Z}_{p^2}} W(\bar{\mathbb{F}}_p)$. Put $\mathbb{L}_z[1/p]=\mathbb{L}_z\otimes_{\mathbb{Z}_{p^2}}\mathbb{Q}_{p^2}$. Then ϕ induces an isomorphism $\phi_* : \mathbb{L}_{z'}[1/p]\xrightarrow{\sim}\mathbb{L}_z[1/p]$. Fix a $\mathbb{Z}_{p^2}$-basis for $\mathbb{L}_z$. Then there exists a $g_{\mathbb{L}}\in \mathrm{GL}_n(\mathbb{Q}_{p^2})$ such that $\phi_*(\mathbb{L}_{z'})=g_{\mathbb{L}}(\mathbb{L}_z)$, and the right coset $g_{\mathbb{L}}\,\mathrm{GL}_n(\mathbb{Z}_{p^2})$ is independent of the choice of such a basis. We put $g_p=(c_0,g_{\mathbb{L}})\in \mathbb{Q}_p^\times\times \mathrm{GL}_n(\mathbb{Q}_{p^2})\simeq G_{0,n}(\mathbb{Q}_p)$, which is well defined up to right multiplication by elements of $K_p=\mathbb{Z}_{p^2}^\times\times\mathrm{GL}_n(\mathbb{Z}_{p^2})$.

Finally, note that the quasi-isogeny $\phi' : B'\to B$ is well determined by z' up to left composition with an element $\gamma\in I(\mathbb{Q})=G_{0,n}(\mathbb{Q})$. If we replace ϕ by $\gamma\circ\phi$, then $g:=(g^p,g_p)\in G_{0,n}(\mathbb{A}^\infty)$ is replaced by $\gamma g=(\gamma g^p,\gamma g_p)$. Therefore, the map

$$\Theta_z : \mathrm{Isog}(z)\to G_{0,n}(\mathbb{Q})\backslash G_{0,n}(\mathbb{A}^\infty)/K,\quad z'\mapsto G_{0,n}(\mathbb{Q})gK$$

is well defined. The fact that Θ_z is a bijection follows from the similar classical statement in characteristic 0. □

Remark 4.12. It follows from Proposition 4.9 and the description of $\mathcal{S}h_{0,n}(\mathbb{C})$ in Section 2.3 that $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ consists of $\#\ker^1(\mathbb{Q},G_{0,n})$ isogeny classes of abelian varieties equipped with additional structures.

Lemma 4.13. *Let N be a fixed nonnegative integer. Up to replacing K^p by an open compact subgroup of itself, the following properties are satisfied: if (B,λ,η) is an $\bar{\mathbb{F}}_p$-point of* $\mathrm{Sh}_{0,n}$ *and $f : B\to B$ is an $\mathcal{O}_D$-quasi-isogeny such that $p^Nf\in$* $\mathrm{End}_{\mathcal{O}_D}(B)$, *$f^\vee\circ\lambda\circ f=\lambda$ and $f\circ\eta=\eta$, then $f=\mathrm{id}$.*

Proof. It suffices to prove the lemma for (B,λ,η) in a fixed isogeny class $\mathrm{Isog}(z)$ of $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. We write $G_{0,n}(\mathbb{A}^\infty)=\coprod_{i\in I}G_{0,n}(\mathbb{Q})g_iK$ with $K=K^pK_p$, where $g_i=g_i^pg_{i,p}$, with $g_i^p\in G_{0,n}(\mathbb{A}^{\infty,p})$ and $g_{i,p}\in G_{0,n}(\mathbb{Q}_p)$, runs through a finite set of representatives of the double coset

$$G_{0,n}(\mathbb{Q})\backslash G_{0,n}(\mathbb{A}^\infty)/K.$$

Let (B,λ,η) be a point of $\mathrm{Sh}_{0,n}$ corresponding to $G_{0,n}(\mathbb{Q})g_iK$ for some $i\in I$, and f be an $\mathcal{O}_D$-quasi-isogeny of B as in the statement. Then f is given by an element of $G_{0,n}^1(\mathbb{Q})$. The condition that $f\circ\eta=\eta$ is equivalent to saying that the image of f in $G_{0,n}(\mathbb{A}^{\infty,p})$ lies in $g_i^pK^pg_i{}^{p,-1}$. Moreover, $p^Nf\in\mathrm{End}_{\mathcal{O}_D}(B)$ implies that the image of f in $G_{0,n}(\mathbb{Q}_p)$ belongs to $\coprod_\delta g_{i,p}(K_p\delta K_p)g_{i,p}^{-1}$, where δ runs through the set

$$\bigl\{(1, \mathrm{Diag}(p^{a_1}, p^{a_2}, \dots, p^{a_n})) \in G_{0,n}(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(\mathbb{Q}_{p^2}) \bigm| 0 \geq a_1 \geq a_2 \geq \cdots \geq a_n \geq -N\bigr\}.$$

Write $\coprod_\delta K_p\delta K_p = \coprod_{j\in J} h_jK_p$ for some finite set J. Hence, it suffices to show that there exists an open compact subgroup $K'^p \subseteq K^p$ such that for all g_i,

$$G^1_{0,n}(\mathbb{Q}) \cap g_i(K'^p \cdot h_jK_p)g_i^{-1} = \{1\}$$

if $h_jK_p = K_p$, and empty otherwise. Since K is neat, we have

$$G^1_{0,n}(\mathbb{Q}) \cap g_i(K'^pK_p)g_i^{-1} = \{1\} \quad \text{for any } g_i \text{ and any } K'^p \subseteq K^p.$$

Note that this implies that, for each $i \in I$, $G^1_{0,n}(\mathbb{Q}) \cap g_i(K^p \cdot h_jK_p)g_i^{-1}$ contains at most one element (because if it contains both x and y, then $x^{-1}y$ is contained in $G^1_{0,n}(\mathbb{Q}) \cap g_iKg_i^{-1} = \{1\}$). Let $S \subset I \times J$ be the subset consisting of (i, j) such that $h_jK_p \neq K_p$ and $G^1_{0,n}(\mathbb{Q}) \cap g_i(K^p \cdot h_jK_p)g_i^{-1}$ indeed contains one element, say $x_{i,j}$. Then $x_{i,j} \neq 1$ for all $(i, j) \in S$. Hence, one can choose a normal open compact subgroup $K'^p \subseteq K^p$ so that $x_{i,j} \notin g_i^pK'^pg_i^{p,-1}$ for all i. We claim that this choice of K'^p will satisfy the desired property. Indeed, if $K^p = \coprod_l b_lK'^p$, then the double coset $G_{0,n}(\mathbb{Q}) \setminus G_{0,n}(\mathbb{A}^\infty)/K'^pK_p$ has a set of representatives of the form g_ib_l. Here, by abuse of notation, we consider b_l as an element of K with p-component equal to 1. Then one has, for $h_jK_p \neq K_p$,

$$G^1_{0,n}(\mathbb{Q}) \cap g_ib_l(K'^ph_jK_p)b_l^{-1}g_i^{-1} = G^1_{0,n}(\mathbb{Q}) \cap g_i(K'^ph_jK_p)g_i^{-1} = \varnothing.$$

The first equality uses the fact that K'^p is normal in K^p. This finishes the proof. □

We come back to the discussion on the cycles $Y_j \subseteq \mathrm{Sh}_{1,n-1}$ for $1 \leq j \leq n$.

Proposition 4.14. *Let $(\mathcal{A}, \lambda, \eta, \mathcal{B}, \lambda', \eta', \phi^{\mathrm{univ}})$ denote the universal object on Y_j for $1 \leq j \leq n$, and $\mathcal{H}_i \subset H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})$ for $i = 1, 2$ be the universal subbundles on $Y'_j \cong Y_j$.*

(1) *The induced map $T_{Y_j} \to \mathrm{pr}_j^* T_{\mathrm{Sh}_{1,n-1}}$ is universally injective, and we have canonical isomorphisms*

$$\begin{aligned} N_{Y_j}(\mathrm{Sh}_{1,n-1}) &:= \mathrm{pr}_j^* T_{\mathrm{Sh}_{1,n-1}}/T_{Y_j} \\ &\cong \bigl(\mathcal{H}_1/V^{-1}(\mathcal{H}_2^{(p)})\bigr)^* \otimes V^{-1}(\mathcal{H}_2^{(p)}) \\ &\qquad \oplus \bigl(F(\mathcal{H}_1^{(p)})/\mathcal{H}_2\bigr) \otimes \bigl(H_1^{\mathrm{dR}}(\mathcal{B}/\mathrm{Sh}_{0,n})_2^\circ/F(\mathcal{H}_1^{(p)})\bigr)^* \\ &\cong \mathrm{Lie}^\circ_{\mathcal{A}^\vee,1} \otimes \mathrm{Coker}(\phi^{\mathrm{univ}}_{*,1}) \oplus \mathrm{Lie}^\circ_{\mathcal{A},2} \otimes \mathrm{Im}(\phi^{\mathrm{univ}}_{*,2})^*. \end{aligned}$$

(2) *Assume that K^p is sufficiently small so that the consequences of Lemma 4.13 hold for $N = 1$. For each fixed closed point $z \in \mathrm{Sh}_{0,n}$, the map $\mathrm{pr}_{j,z} := \mathrm{pr}_j|_{Y_{j,z}} : Y_{j,z} \to \mathrm{Sh}_{1,n-1}$ is a closed immersion, or equivalently, the morphism $(\mathrm{pr}_j, \mathrm{pr}'_j) : Y_j \to \mathrm{Sh}_{1,n-1} \times_{\mathrm{Spec}(\mathbb{F}_{p^2})} \mathrm{Sh}_{0,n}$ is a closed immersion.*

(3) *The union of the images of pr_j for all $1 \le j \le n$ is the supersingular locus of $\mathrm{Sh}_{1,n-1}$, i.e., the reduced closed subscheme of $\mathrm{Sh}_{1,n-1}$ where all the slopes of the Newton polygon of the p-divisible group $\mathcal{A}[\mathfrak{p}^\infty]$ are $1/2$.*

Proof. **(1)** Let S be an affine noetherian $\mathbb{F}_{p^2}$-scheme and let $y = (A, \lambda, \eta, B, \lambda', \eta', \phi)$ be an S-point of Y_j. Put $\hat{S} = S \times_{\mathrm{Spec}(\mathbb{F}_{p^2})} \mathrm{Spec}(\mathbb{F}_{p^2}[t]/t^2)$. Then we have a natural bijection

$$\mathscr{D}\mathrm{ef}(y, \hat{S}) \cong \Gamma(S, y^* T_{Y_j}),$$

where $\mathscr{D}\mathrm{ef}(y, \hat{S})$ is the set of deformations of y to $\hat{S}$. Similarly, $\mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S}) \cong \Gamma(S, y^* \mathrm{pr}_j^* T_{\mathrm{Sh}_{1,n-1}})$. To prove the universal injectivity of $T_{Y_j} \to \mathrm{pr}_j^* T_{\mathrm{Sh}_{1,n-1}}$, it suffices to show that the natural map $\mathscr{D}\mathrm{ef}(y, \hat{S}) \to \mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S})$ is injective. By crystalline deformation theory (Theorem 3.4), giving a point of $\mathscr{D}\mathrm{ef}(y, \hat{S})$ is equivalent to giving $\mathcal{O}_{\hat{S}}$-subbundles $\hat{\omega}^\circ_{A^\vee,i} \subseteq H_1^{\mathrm{cris}}(A/\hat{S})_i^\circ$ over $\hat{S}$ for $i = 1, 2$ such that

- $\hat{\omega}^\circ_{A^\vee,i}$ lifts $\omega^\circ_{A^\vee/S,i}$;
- $\hat{\omega}^\circ_{A^\vee,1} \subseteq \mathrm{Im}(\phi_{*,1}) \otimes \mathbb{F}_{p^2}[t]/t^2$ and $\mathrm{Im}(\phi_{*,2}) \otimes \mathbb{F}_{p^2}[t]/t^2 \subseteq \hat{\omega}^\circ_{A^\vee,2}$ are locally direct factors.

Hence, one sees easily that

$$\begin{aligned}\mathscr{D}\mathrm{ef}(y, \hat{S}) &\cong \mathrm{Hom}_{\mathcal{O}_S}(\omega^\circ_{A^\vee/S,1}, \mathrm{Im}(\phi_{*,1})/\omega^\circ_{A^\vee/S,1}) \\ &\qquad \oplus \mathrm{Hom}_{\mathcal{O}_S}(\omega^\circ_{A^\vee/S,2}/\mathrm{Im}(\phi_{*,2}), H_1^{\mathrm{dR}}(A/S)_2^\circ/\omega^\circ_{A^\vee/S,2}) \\ &\cong \mathrm{Lie}^\circ_{A^\vee/S,1} \otimes (\mathrm{Im}(\phi_{*,1})/\omega^\circ_{A^\vee/S,1}) \oplus (\omega^\circ_{A^\vee/S,2}/\mathrm{Im}(\phi_{*,2}))^* \otimes \mathrm{Lie}^\circ_{A/S,2}.\end{aligned}$$

Similarly, $\mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S})$ is given by the lifts of $\omega^\circ_{A^\vee/S,i}$ to $\hat{S}$ for $i = 1, 2$. These lifts are classified by $\mathrm{Hom}_{\mathcal{O}_S}(\omega^\circ_{A^\vee/S,i}, H_1^{\mathrm{dR}}(A/S)_i^\circ/\omega^\circ_{A^\vee/S,i}) \cong \mathrm{Lie}^\circ_{A^\vee/S,i} \otimes_k \mathrm{Lie}^\circ_{A/S,i}$. Hence, $\mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S})$ is canonically isomorphic to

$$\mathrm{Lie}^\circ_{A^\vee/S,1} \otimes_{\mathcal{O}_S} \mathrm{Lie}^\circ_{A/S,1} \oplus \mathrm{Lie}^\circ_{A^\vee/S,2} \otimes_{\mathcal{O}_S} \mathrm{Lie}^\circ_{A/S,2}.$$

The natural map $\mathscr{D}\mathrm{ef}(y, \hat{S}) \to \mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S})$ is induced by the natural maps

$$\begin{aligned}\mathrm{Im}(\phi_{*,1})/\omega^\circ_{A^\vee/S,1} &\hookrightarrow H_1^{\mathrm{dR}}(A/S)_1^\circ/\omega^\circ_{A^\vee/S,1} \cong \mathrm{Lie}^\circ_{A/S,1}, \\ (\omega^\circ_{A^\vee/S,2}/\mathrm{Im}(\phi_{*,2}))^* &\hookrightarrow \qquad \omega^{\circ,*}_{A^\vee/S,2} \qquad \cong \mathrm{Lie}^\circ_{A^\vee/S,2}.\end{aligned}$$

It follows that $\mathscr{D}\mathrm{ef}(y, \hat{S}) \to \mathscr{D}\mathrm{ef}(\mathrm{pr}_j \circ y, \hat{S})$ is injective. To prove the formula for $N_{Y_j}(\mathrm{Sh}_{1,n-1})$, we apply the arguments above to affine open subsets of Y_j. We see easily that

$$\begin{aligned}N_{Y_j}(\mathrm{Sh}_{1,n-1}) &\cong \mathrm{Lie}^{\circ}_{\mathcal{A}^\vee,1}\otimes_{\mathcal{O}_{Y_j}}\mathrm{Coker}(\phi^{\mathrm{univ}}_{*,1})\oplus \mathrm{Lie}^{\circ}_{\mathcal{A},2}\otimes_{\mathcal{O}_{Y_j}}\mathrm{Im}(\phi^{\mathrm{univ}}_{*,2})^*\\ &\cong \big(\mathcal{H}_1/V^{-1}(\mathcal{H}_2^{(p)})\big)^*\otimes V^{-1}(\mathcal{H}_2^{(p)})\\ &\qquad \oplus \big(F(\mathcal{H}_1^{(p)})/\mathcal{H}_2\big)\otimes \big(H_1^{\mathrm{dR}}(\mathcal{B}/Y_j)^{\circ}_2/F(\mathcal{H}_1^{(p)})\big)^*.\end{aligned}$$

Here, the last step uses (4.7.2) and the isomorphism

$$\mathrm{Im}(\phi^{\mathrm{univ}}_{*,2})\cong H_1^{\mathrm{dR}}(\mathcal{B}/Y_j)^{\circ}_2/\operatorname{Ker}(\phi^{\mathrm{univ}}_{*,2})\cong H_1^{\mathrm{dR}}(\mathcal{B}/Y_j)^{\circ}_2/F(\mathcal{H}_1^{(p)}).$$

(2) By statement (1), $\mathrm{pr}_{j,z}$ induces an injection of tangent spaces at each closed points of $Y_{j,z}$. To complete the proof, it suffices to prove that $\pi_{j,z}$ induces injections on the closed points. Write $z=(B,\lambda',\eta')\in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. Assume y_1 and y_2 are two closed points of $Y_{j,z}$ with $\pi_j(y_1)=\pi_j(y_2)=(A,\lambda,\eta)$. Let $\phi_1,\phi_2:B\to A$ be the isogenies given by y_1 and y_2. Then the quasi-isogeny $\phi_{1,2}=\phi_2^{-1}\phi_1\in \mathrm{End}_{\mathcal{O}_D}(B)_{\mathbb{Q}}$ satisfies the conditions of Lemma 4.13 for $N=1$. Hence, we get $\phi_{1,2}=\mathrm{id}_B$, which is equivalent to $y_1=y_2$. This proves that $\pi_{j,z}$ is injective on closed points.

(3) The proof resembles the work of Vollaard and Wedhorn [2011]. Since all the points of $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ are supersingular by Remark 3.7, it is clear that the image of each pr_j lies in the supersingular locus of $\mathrm{Sh}_{1,n-1}$. Suppose now we are given a supersingular point $x=(A,\lambda,\eta)\in \mathrm{Sh}_{1,n-1}(\bar{\mathbb{F}}_p)$. We have to show that there exists $(B,\lambda',\eta')\in \mathrm{Sh}_{0,n}$ and an isogeny $\phi:B\to A$ such that $(A,\lambda,\eta,\lambda',\eta';\phi)$ lies in Y_j for some $1\le j\le n$.

Consider

$$\mathbb{L}_{\mathbb{Q}}=(\tilde{\mathcal{D}}(A)^{\circ}_1[1/p])^{F^2=p}=\{a\in \tilde{\mathcal{D}}(A)^{\circ}_1[1/p]\mid F^2(a)=pa\}.$$

Since x is supersingular, $\mathbb{L}_{\mathbb{Q}}$ is a $\mathbb{Q}_{p^2}$-vector space of dimension n by the Dieudonné–Manin classification, and $\tilde{\mathcal{D}}(A)^{\circ}_1[1/p]=\mathbb{L}_{\mathbb{Q}}\otimes_{\mathbb{Q}_{p^2}}W(\bar{\mathbb{F}}_p)[1/p]$. We put $\tilde{\mathcal{E}}^{\circ}_1=(\mathbb{L}_{\mathbb{Q}}\cap\tilde{\mathcal{D}}(A)^{\circ}_1)\otimes_{\mathbb{Z}_{p^2}}W(\bar{\mathbb{F}}_p)$, and $\tilde{\mathcal{E}}^{\circ}_2=F(\tilde{\mathcal{E}}^{\circ}_1)\subseteq\tilde{\mathcal{D}}(A)^{\circ}_2$. Thus $\tilde{\mathcal{E}}^{\circ}=\tilde{\mathcal{E}}^{\circ}_1\oplus\tilde{\mathcal{E}}^{\circ}_2$ is a Dieudonné submodule of $\tilde{\mathcal{D}}(A)^{\circ}$. We claim that $\tilde{\mathcal{E}}^{\circ}$ contains $p\tilde{\mathcal{D}}(A)^{\circ}$ as a submodule. Then applying Proposition 3.2 with $m=1$, we get an $\mathcal{O}_D$-abelian variety (B,λ',η') together with an $\mathcal{O}_D$-isogeny $\phi:B\to A$ with $\phi^\vee\circ\lambda\circ\phi=p\lambda$. It is easy to see in this case that $(A,\lambda,\eta,B,\lambda',\eta',\phi)$ defines a point in Y_j with $j=\dim_{\bar{\mathbb{F}}_p}(\tilde{\mathcal{D}}(A)^{\circ}_2/\tilde{\mathcal{E}}^{\circ}_2)$.

It then suffices to prove the claim that $p\tilde{\mathcal{D}}(A)^{\circ}\subseteq\tilde{\mathcal{E}}^{\circ}$. Suppose not, then $\tilde{\mathcal{D}}(A)^{\circ}\nsubseteq(1/p)\tilde{\mathcal{E}}^{\circ}$. Consider $M_i:=\tilde{\mathcal{D}}(A)^{\circ}_i/\tilde{\mathcal{E}}^{\circ}_i$ for $i=1,2$. For any integer $\alpha\ge 0$, its p^α-torsion submodule is

$$M_i[p^\alpha]=\Big(\tilde{\mathcal{D}}(A)^{\circ}_i\cap\frac{1}{p^\alpha}\tilde{\mathcal{E}}^{\circ}_i\Big)/\tilde{\mathcal{E}}^{\circ}_i.$$

It follows easily that

$$M_i[p^{\alpha+1}]/M_i[p^\alpha]\cong\Big(\frac{1}{p^{\alpha+1}}\tilde{\mathcal{E}}^{\circ}_i\cap\Big(\tilde{\mathcal{D}}(A)^{\circ}_i+\frac{1}{p^\alpha}\tilde{\mathcal{E}}^{\circ}_i\Big)\Big)\Big/\frac{1}{p^\alpha}\tilde{\mathcal{E}}^{\circ}_i.$$

On the other hand, the Kottwitz's signature condition implies that both F and $V : \tilde{\mathcal{D}}(A)^\circ_1 \to \tilde{\mathcal{D}}(A)^\circ_2$ have cokernel isomorphic to $\bar{\mathbb{F}}_p$, and both F and $V : \tilde{\mathcal{E}}^\circ_1 \to \tilde{\mathcal{E}}^\circ_2$ are isomorphism. Therefore, the two induced morphisms

$$F \text{ and } V : M_1 \to M_2$$

are injective and both have cokernel isomorphic to $\bar{\mathbb{F}}_p$. It follows that the induced maps on the graded pieces

$$\begin{aligned} F \text{ and } V : \Big(\frac{1}{p^{\alpha+1}}\tilde{\mathcal{E}}^\circ_1 \cap \Big(\tilde{\mathcal{D}}(A)^\circ_1 + \frac{1}{p^\alpha}\tilde{\mathcal{E}}^\circ_1\Big)\Big)\Big/\frac{1}{p^\alpha}\tilde{\mathcal{E}}^\circ_1 \\ \to \Big(\frac{1}{p^{\alpha+1}}\tilde{\mathcal{E}}^\circ_2 \cap \Big(\tilde{\mathcal{D}}(A)^\circ_2 + \frac{1}{p^\alpha}\tilde{\mathcal{E}}^\circ_2\Big)\Big)\Big/\frac{1}{p^\alpha}\tilde{\mathcal{E}}^\circ_2 \end{aligned} \tag{4.14.1}$$

are injective maps, and are isomorphisms for all $\alpha \geq 0$ except for exactly one α.[17] The assumption $\tilde{\mathcal{D}}(A)^\circ \nsubseteq (1/p)\tilde{\mathcal{E}}^\circ$ implies that there are at least two $\alpha \geq 0$ for which the right hand side of (4.14.1) is nonzero. So there exists $\alpha \geq 0$ such that (4.14.1) are isomorphisms of *nonzero* $\bar{\mathbb{F}}_p$*-vector spaces*. Multiplication by p^α gives isomorphisms:

$$F \text{ and } V : \Big(\frac{1}{p}\tilde{\mathcal{E}}^\circ_1 \cap \Big(p^\alpha\tilde{\mathcal{D}}(A)^\circ_1 + \tilde{\mathcal{E}}^\circ_1\Big)\Big) \to \Big(\frac{1}{p}\tilde{\mathcal{E}}^\circ_2 \cap \Big(p^\alpha\tilde{\mathcal{D}}(A)^\circ_2 + \tilde{\mathcal{E}}^\circ_2\Big)\Big). \tag{4.14.2}$$

Now, Hilbert 90 theorem implies that $\mathbb{L}' = \big((1/p)\tilde{\mathcal{E}}^\circ_1 \cap (p^\alpha\tilde{\mathcal{D}}(A)^\circ_1 + \tilde{\mathcal{E}}^\circ_1)\big)^{F^2=p}$ in fact generates the source of (4.14.2). On the other hand, it is obvious that $\mathbb{L}' \subset \mathbb{L}_{\mathbb{Q}}$ and $\mathbb{L}' \subseteq p^\alpha\tilde{\mathcal{D}}(A)^\circ_1 + \tilde{\mathcal{E}}^\circ_1 \subseteq \tilde{\mathcal{D}}(A)^\circ_1$. This means that $\mathbb{L}'$, and hence $\mathbb{L}_{\mathbb{Q}} \cap \tilde{\mathcal{D}}(A)^\circ_1$, generates the entire $(1/p)\tilde{\mathcal{E}}^\circ_1 \cap (p^\alpha\tilde{\mathcal{D}}(A)^\circ_1 + \tilde{\mathcal{E}}^\circ_1)$, i.e., one has $(1/p)\tilde{\mathcal{E}}^\circ_1 \cap (p^\alpha\tilde{\mathcal{D}}(A)^\circ_1 + \tilde{\mathcal{E}}^\circ_1) = \tilde{\mathcal{E}}^\circ_1$. But this contradicts with the nontriviality of the vector spaces in (4.14.1) by our choice of α. Now the proposition is proved. □

Corollary 4.15. *The morphism* pr_1 (*resp.* pr_n) *is a closed immersion*, *and it identifies* Y_1 (*resp.* Y_n) *with the closed subscheme of* $\mathrm{Sh}_{1,n-1}$ *defined by the vanishing of* $V : \omega^\circ_{\mathcal{A}^\vee,2} \to \omega^{\circ,(p)}_{\mathcal{A}^\vee,1}$ (*resp.* $V : \omega^\circ_{\mathcal{A}^\vee,1} \to \omega^{\circ,(p)}_{\mathcal{A}^\vee,2}$).

Proof. We just prove the statement for pr_1, and the case of pr_n is similar. Let Z_1 be the closed subscheme of $\mathrm{Sh}_{1,n-1}$ defined by the condition that $V : \omega^\circ_{\mathcal{A}^\vee,2} \to \omega^{\circ,(p)}_{\mathcal{A}^\vee,1}$ vanishes. We show first that $\mathrm{pr}_1 : Y_1 \to \mathrm{Sh}_{1,n-1}$ factors through the natural inclusion $Z_1 \hookrightarrow \mathrm{Sh}_{1,n-1}$. Let $y = (A, \lambda, \eta, B, \lambda', \eta', \phi)$ be an S-valued point of Y_1. By Lemma 4.6, $\mathrm{Im}(\phi_{2,*})$ has rank $n-1$ and contains both $\omega^\circ_{A^\vee/S,2}$ and $F(H_1^{\mathrm{dR}}(A/S)_1^{\circ,(p)})$, which are both $\mathcal{O}_S$-subbundles of rank $n-1$. This forces $\omega^\circ_{A^\vee/S,2} = F(H_1^{\mathrm{dR}}(A/S)_1^{\circ,(p)})$, and therefore $V : \omega^\circ_{A^\vee/S,2} \to \omega^{\circ,(p)}_{A^\vee/S,1}$ vanishes. This shows that $\mathrm{pr}_1(y) \in Z_1$.

[17] We point out that, for (4.14.1), F is an isomorphism if and only if V is an isomorphism, because this is equivalent to requiring the source and the target to have the same dimension.

To prove that $\mathrm{pr}_1 : Y_1 \to Z_1$ is an isomorphism, as Y_1 is smooth, it suffices to show that it induces a bijection between closed points and tangent spaces of Y_1 and Z_1. For any perfect field k containing $\mathbb{F}_{p^2}$, one constructs a map $\theta : Z_1(k) \to Y_1(k)$ inverse to $\mathrm{pr}_1 : Y_1(k) \to Z_1(k)$ as follows. Given $x = (A, \lambda, \eta) \in Z_1(k)$. Let $\tilde{\mathcal{E}}_1^\circ = \tilde{\mathcal{D}}(A)_1^\circ$ and $\tilde{\mathcal{E}}_2^\circ \subseteq \tilde{\mathcal{D}}(A)_2^\circ$ be the inverse image of $\omega^\circ_{A^\vee/k,2} \subseteq \tilde{\mathcal{D}}(A)_2^\circ / p\tilde{\mathcal{D}}(A)_2^\circ$. Then the condition that $y \in Z_1$ implies that $\tilde{\mathcal{E}}_1^\circ \oplus \tilde{\mathcal{E}}_2^\circ$ is stable under F and V. Applying Proposition 3.2 with $m = 1$, we get a tuple $(B, \lambda', \eta', \phi)$ such that $y = (A, \lambda, \eta, B, \lambda', \eta', \phi) \in Y_1(k)$. It is immediate to check that $x \mapsto y$ and pr_1 are the set theoretic inverse of each other. It remains to show that pr_1 induces a bijection between $T_{Y_1,y}$ and $T_{Z_1,x}$. Proposition 4.14 already implies that we have an inclusion $T_{Y_1,y} \hookrightarrow T_{Z_1,x} \hookrightarrow T_{\mathrm{Sh}_{1,n-1},x}$. It suffices to check that $\dim T_{Z_1,x} = n-1$. The tangent space $T_{Z_1,x}$ is the space of deformations $(\hat{A}, \hat{\lambda}, \hat{\eta})$ over $\hat{k} = k[\epsilon]/(\epsilon^2)$ of (A, λ, η) such that $V : \omega^\circ_{\hat{A}^\vee/\hat{k},2} \to \omega^{\circ,(p)}_{\hat{A}^\vee/\hat{k},1} = \omega^{\circ,(p)}_{A^\vee/k,1} \otimes_k \hat{k}$ vanishes. This uniquely determines the lift $\hat{\omega}^\circ_{A^\vee,2} = \omega^\circ_{\hat{A}^\vee/\hat{k},2}$. So by deformation theory (Theorem 3.4), the tangent space $T_{Z_1,x}$ is determined by the liftings $\hat{\omega}^\circ_{A^\vee,1} = \omega^\circ_{\hat{A}^\vee/\hat{k},1}$ of $\omega^\circ_{A^\vee/k,1}$. So it is of dimension $n-1$. This concludes the proof of the corollary. □

4.16. ***Geometric Jacquet–Langlands morphism.*** Let $\ell \neq p$ be a prime number. For $1 \le j \le n$, the diagram (4.2.1) gives rise to a natural morphism

$$\mathcal{JL}_j : H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell) \xrightarrow{\mathrm{pr}'^*_j} H^0_{\mathrm{et}}(\bar{Y}_j, \overline{\mathbb{Q}}_\ell) \xrightarrow{\mathrm{pr}_{j,!}} H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1)), \quad (4.16.1)$$

where $\mathrm{pr}_{j,!}$ is (2.10.1), whose restriction to each $H^0_{\mathrm{et}}(Y_{j,z}, \overline{\mathbb{Q}}_\ell)$ for $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ is the Gysin map associated to the closed immersion $Y_{j,z} \hookrightarrow \overline{\mathrm{Sh}}_{1,n-1}$. It is clear that the image of $\mathcal{JL}_j$ is the subspace generated by the cycle classes of $[Y_{j,z}] \in A^{n-1}(\overline{\mathrm{Sh}}_{1,n-1})$ with $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. According to [Helm 2010], $\mathcal{JL}_j$ should be considered as a certain geometric realization of the Jacquet–Langlands transfer from $G_{0,n}$ to $G_{1,n-1}$. Putting all the $\mathcal{JL}_j$ together, we get a morphism

$$\mathcal{JL} = \sum_j \mathcal{JL}_j : \bigoplus_{j=1}^n H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell) \to H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1)). \quad (4.16.2)$$

Recall that we have fixed an isomorphism $G_{1,n-1}(\mathbb{A}^\infty) \simeq G_{0,n}(\mathbb{A}^\infty)$, which we write uniformly as $G(\mathbb{A}^\infty)$. Denote by $\mathscr{H}(K^p, \overline{\mathbb{Q}}_\ell) = \overline{\mathbb{Q}}_\ell[K^p \backslash G(\mathbb{A}^{\infty,p})/K^p]$ the prime-to-p Hecke algebra. Then the homomorphism (4.16.2) is a homomorphism of $\mathscr{H}(K^p, \overline{\mathbb{Q}}_\ell)$-modules.

For an irreducible admissible representation π of $G(\mathbb{A}_\infty)$, we write $\pi = \pi^p \otimes \pi_p$, where π^p (resp. π_p) is the prime-to-p part (resp. the p-component) of π.

Lemma 4.17. *Let π_1 and π_2 be two admissible irreducible representations of $G(\mathbb{A}^\infty)$, and (r_i, s_i) for $i = 1, 2$ be two pairs of integers with $0 \le r_i, s_i \le n$ and $r_1 + s_1 \equiv r_2 + s_2 \mod 2$. Assume that π_1 satisfies Hypothesis 2.5 with $a_\bullet = (r_1, s_1)$,* *and there exists an admissible irreducible representation* $\pi_{2,\infty}$ *of* $G_{(r_2,s_2)}(\mathbb{R})$ *such that* $\pi_2 \otimes \pi_{2,\infty}$ *is a cuspidal automorphic representation of* $G_{(r_2,s_2)}(\mathbb{A})$. *If* π_1^p *and* π_2^p *are isomorphic as representations of* $G(\mathbb{A}^{p,\infty})$, *then* $\pi_{1,p} \simeq \pi_{2,p}$, *and* $\pi_2 \otimes \pi_{2,\infty}$ *admits a base change to a cuspidal automorphic representation of* $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$; *in particular*, π_2 *satisfies Hypothesis 2.5 for* $a_\bullet = (r_2, s_2)$.

Proof. By assumption on π_1, there exists an irreducible admissible representation $\pi_{1,\infty}$ of $G_{(r_1,s_1)}(\mathbb{R})$ such that $\pi_1 \otimes \pi_{1,\infty}$ is a cuspidal automorphic representation of $G_{r_1,s_1}(\mathbb{A})$, which base changes to a cuspidal automorphic representation (Π_1, χ_1) of $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$. On the other hand, by [Shin 2014, Theorem 1.1], there exists always a base change of $\pi_2 \otimes \pi_{2,\infty}$ to an automorphic representation (Π_2, χ_2) of $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$. The base changes (Π_i, χ_i) with $i = 1, 2$ satisfy the following properties:

- Π_i is conjugate self-dual,
- for every unramified rational prime x, the x-component of (Π_i, ψ_i) depends only on the x-component of π_i and
- if $\pi_{i,p} = \pi_{i,0} \otimes \pi_{i,\mathfrak{p}}$ as representation of $G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_\mathfrak{p})$, then $\Pi_{i,p} = (\pi_{i,\mathfrak{p}} \otimes \check{\pi}_{i,\mathfrak{p}}^c)$ as a representation of $\mathrm{GL}_n(E \otimes \mathbb{Q}_p) \cong \mathrm{GL}_n(E_\mathfrak{p}) \times \mathrm{GL}_n(E_{\bar{\mathfrak{p}}})$, and $\psi_{i,p} = \pi_{i,0} \otimes \pi_{i,0}^{-1}$ as a representation of $(E_0 \otimes \mathbb{Q}_p)^\times = \mathbb{Q}_p^\times \times \mathbb{Q}_p^\times$. Here, $\check{\pi}_{i,\mathfrak{p}}^c$ denotes the complex conjugate of the contragredient of $\pi_{i,\mathfrak{p}}$.

As $\pi_1^p \simeq \pi_2^p$, (Π_1, ψ_1) and (Π_2, ψ_2) are isomorphic at almost all finite places. By the strong multiplicity one theorem for GL_n [Jacquet and Shalika 1981], we have $(\Pi_1, \psi_1) \simeq (\Pi_2, \psi_2)$; in particular, (Π_2, ψ_2) is cuspidal. By the description of $(\Pi_{i,p}, \psi_{i,p})$, it follows immediately that $\pi_{1,p} \simeq \pi_{2,p}$. □

Let $\mathcal{A}_K$ be the set of isomorphism classes of irreducible admissible representations π of $G(\mathbb{A}^\infty)$ satisfying Hypothesis 2.5 with $a_\bullet = (0, n)$. In particular, each $\pi \in \mathcal{A}_K$ is the finite part of an automorphic cuspidal representation of $G_{0,n}(\mathbb{A})$.

We fix such a $\pi \in \mathcal{A}_K$. Let

$$\mathcal{JL}_\pi : \bigoplus_{i=1}^{n} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_{\pi^p} \to H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))_{\pi^p}$$

denote the homomorphism on the $(\pi^p)^{K^p}$-isotypic components induced by $\mathcal{JL}$, where for an $\mathcal{H}(K^p, \overline{\mathbb{Q}}_\ell)$-module M we put

$$M_{\pi^p} := \mathrm{Hom}_{\mathcal{H}(K^p,\overline{\mathbb{Q}}_\ell)}((\pi^p)^{K^p}, M) \otimes (\pi^p)^{K^p}.$$

Then Lemma 4.17 implies that π is completely determined by its prime-to-p part. Hence, taking the π^p-isotypic components is the same as taking the π-isotypic components. We can thus write M_π instead of M_{π^p} for a $\mathcal{H}(K, \overline{\mathbb{Q}}_\ell)$-module M.

Recall that the image of $\mathcal{JL}_\pi$ is included in $H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi$, which is the maximal subspace of $H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))_\pi$ where the action of $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^2})$ factors through a finite quotient. Note that, at this moment, it is not clear if the target of $\mathcal{JL}_\pi$ is nonzero. But this will follow from the proof of our main Theorem 4.18 below.

Our main result claims that this inclusion is actually an equality under certain genericity conditions on π_p. To make this precise, write $\pi_p = \pi_{p,0} \otimes \pi_{\mathfrak{p}}$ as a representation of $G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_{\mathfrak{p}})$. Let

$$\rho_{\pi_{\mathfrak{p}}} : W_{\mathbb{Q}_{p^2}} \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$$

be the unramified representation of the Weil group of $\mathbb{Q}_{p^2}$ defined in (2.6.1). It induces a continuous ℓ-adic representation of $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^2})$, which we denote by the same notation. Then $\rho_{\pi_{\mathfrak{p}}}(\mathrm{Frob}_{p^2})$ is semisimple with characteristic polynomial (2.6.2). Using this, we get an explicit description of $H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))_\pi$ and $H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_\pi$ in terms of $\rho_{\pi_{\mathfrak{p}}}$ by (2.4.1) and (2.6.3).

We can now state our main theorem.

Theorem 4.18. *Fix a π in $\mathscr{A}_K$. Let $\alpha_{\pi_{\mathfrak{p}},1}, \dots, \alpha_{\pi_{\mathfrak{p}},n}$ be the eigenvalues of $\rho_{\pi_{\mathfrak{p}}}(\mathrm{Frob}_{p^2})$.*

(1) *If $\alpha_{\pi_{\mathfrak{p}},1}, \dots, \alpha_{\pi_{\mathfrak{p}},n}$ are distinct, then the map $\mathcal{JL}_\pi$ is injective;*

(2) *Let $m_{1,n-1}(\pi)$ (resp. $m_{0,n}(\pi)$) denote the multiplicity for π appearing in Theorem 2.6 for $a_\bullet = (1, n-1)$ (resp. for $a_\bullet = (0, n)$). Assume that $m_{1,n-1}(\pi) = m_{0,n}(\pi)$ and that $\alpha_{\pi_{\mathfrak{p}},i}/\alpha_{\pi_{\mathfrak{p}},j}$ is not a root of unity for all $1 \le i, j \le n$. Then the map*

$$\mathcal{JL}_\pi : \bigoplus_{j=1}^{n} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_\pi \to H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi$$

is an isomorphism. In other words, $H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi$ is generated by the cycle classes of the irreducible components of Y_j for $1 \le j \le n$.

The proof of this theorem will be given at the end of Section 6.

Remark 4.19. The equality $m_{1,n-1}(\pi) = m_{0,n}(\pi)$ is conjectured to be true according to Arthur's formula on the automorphic multiplicities of unitary groups, and is known to hold when π is the finite part of an automorphic representation of $G_{1,n-1}(\mathbb{A})$ whose base change to $\mathrm{GL}_n(\mathbb{A}_E) \times \mathbb{A}_{E_0}^\times$ is cuspidal, and $G_{1,n-1}$ is quasisplit at all finite places. See for instance [White 2012, Theorem E].

On the other hand, Theorem 4.18(1) gives partial results towards the equality $m_{1,n-1}(\pi) = m_{0,n}(\pi)$. Indeed, when combining with Kottwitz's description 2.6 of the π-isotypic components of the cohomology groups, Theorem 4.18(1) implies (under the assumption that the Satake parameters of $\pi_{\mathfrak{p}}$ are distinct) that $m_{1,n-1}(\pi) \ge m_{0,n}(\pi)$ without using Arthur's trace formula. If we use only the fact that $\mathcal{JL}_\pi$

is nonzero (which is an easy consequence of our computation of the intersection matrix in Theorem 6.7), we get the implication $m_{0,n}(\pi) \neq 0 \Rightarrow m_{1,n-1}(\pi) \neq 0$.

5. Fundamental intersection numbers

In this section, we will compute some intersection numbers on certain Deligne–Lusztig varieties. These numbers will play a key role in the computation in the next section of the intersection matrix of the cycles Y_j on $\mathrm{Sh}_{1,n-1}$.

Notation 5.1. Let X be an algebraic variety of pure dimension N over $\bar{\mathbb{F}}_p$. For an integer $r \geq 0$, let $A^r(X)$ (resp. $A_r(X)$) denote the group of algebraic cycles on X of codimension r (resp. of dimension r) modulo rational equivalences. If $Y \subseteq X$ is a subscheme equidimensional of codimension r, we denote by $[Y] \in A^r(X)$ the class of Y. We put $A^\star(X) = \bigoplus_{r=0}^N A^r(X)$. For a zero-dimensional cycle $\eta \in A^N(X)$, we denote by

$$\deg(\eta) = \int_X \eta$$

the *degree* of η. Let $\mathcal{V}$ be a vector bundle over X. We denote by $c_r(\mathcal{V}) \in A^r(X)$ the r-th Chern class of $\mathcal{V}$ for $0 \leq r \leq N$, and put $c(\mathcal{V}) = \sum_{r=0}^N c_r(\mathcal{V}) t^r$ in the free variable t.

5.2. ***A special Deligne–Lusztig variety.*** We fix an integer $n \geq 1$. For an integer $0 \leq k \leq n$, we denote by $\boldsymbol{Gr}(n,k)$ the Grassmannian variety over $\mathbb{F}_p$ classifying k-dimensional subspaces of $\mathbb{F}_p^{\oplus n}$. Given an integer k with $1 \leq k \leq n$, let $Z_k^{\langle n\rangle}$ be the subscheme of $\boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1)$ whose S-valued points are isomorphism classes of pairs (L_1, L_2), where L_1 and L_2 are respectively subbundles of $\mathcal{O}_S^{\oplus n}$ of rank k and $k-1$ satisfying $L_2 \subseteq L_1^{(p)}$ and $L_2^{(p)} \subseteq L_1$ (with locally free quotients). The same arguments as in Proposition 4.4 show that $Z_k^{\langle n\rangle}$ is a smooth variety over $\mathbb{F}_p$ of dimension $n-1$. We denote the natural closed immersion by

$$i_k : Z_k^{\langle n\rangle} \hookrightarrow \boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1).$$

Let $\mathcal{L}_1$ and $\mathcal{L}_2$ denote the universal subbundles on $\boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1)$ coming from the two factors, and $\mathcal{Q}_1$ and $\mathcal{Q}_2$ the universal quotients, respectively. When there is no confusion, we still use $\mathcal{L}_i$ and $\mathcal{Q}_i$ for $i = 1, 2$ to denote their restrictions to $Z_k^{\langle n\rangle}$. We put

$$\mathcal{E}_k = (\mathcal{L}_1/\mathcal{L}_2^{(p)})^* \otimes \mathcal{L}_2^{(p)} \oplus (\mathcal{L}_1^{(p)}/\mathcal{L}_2) \otimes \mathcal{Q}_1^{*,(p)}, \tag{5.2.1}$$

which is a vector bundle of rank $n-1$ on $Z_k^{\langle n\rangle}$. (This vector bundle is modeled on the description of the normal bundle $N_{Y_j}(\mathrm{Sh}_{1,n-1})$ in Proposition 4.14(1), which is how our computation will be used in the next section; see Proposition 6.4.) We have the top Chern class $c_{n-1}(\mathcal{E}_k) \in A^{n-1}(Z_k^{\langle n \rangle})$. We define the *fundamental intersection number* on $Z_k^{\langle n \rangle}$ as

$$N(n,k) := \int_{Z_k^{\langle n \rangle}} c_{n-1}(\mathcal{E}_k). \tag{5.2.2}$$

The main theorem we prove in this section is the following:

Theorem 5.3. *For integers n, r with $0 \le r \le n$, let*

$$\binom{n}{r}_q = \frac{(q^n-1)(q^{n-1}-1)\cdots(q^{n-r+1}-1)}{(q-1)(q^2-1)\cdots(q^r-1)}$$

be the Gaussian binomial coefficients, and let $d(n,k) = (2k-1)n - 2k(k-1) - 1$ denote the dimension of $\boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1)$. Then, for $1 \le k \le n$, we have

$$N(n,k) = (-1)^{n-1} \sum_{\delta=0}^{\min\{k-1,n-k\}} (n-2\delta)\, p^{d(n-2\delta,k-\delta)} \binom{n}{\delta}_{p^2}. \tag{5.3.1}$$

Remark 5.4. We point out that this theorem seems to be more than a technical result. It is at the heart of the understanding of these cycles we constructed.

Proof. We first claim that $N(n,k) = N(n, n+1-k)$ for $1 \le k \le n$. Let (L_1, L_2) be an S-valued point of $\boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1)$, and $Q_i = \mathcal{O}_S^{\oplus n}/L_i$ for $i = 1, 2$ be the corresponding quotient bundles. Then $(L_1, L_2) \mapsto (Q_2^*, Q_1^*)$ defines a duality isomorphism

$$\theta : \boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1) \xrightarrow{\sim} \boldsymbol{Gr}(n, n+1-k) \times \boldsymbol{Gr}(n, n-k).$$

Since $L_2^{(p)} \subseteq L_1$ (resp. $L_2 \subseteq L_1^{(p)}$) is equivalent to $Q_1^* \subseteq Q_2^{*,(p)}$ (resp. to $Q_1^{*,(p)} \subseteq Q_2^*$), θ induces an isomorphism between $Z_k^{(n)}$ and $Z_{n+1-k}^{(n)}$. It is also direct to check that $\mathcal{E}_k = \theta^*(\mathcal{E}_{n+1-k})$. This verifies the claim. Now since the right hand side of (5.3.1) is also invariant under replacing k by $n+1-k$, it suffices to prove the theorem when $k \le \frac{1}{2}(n+1)$.

We reduce the proof of the theorem to an analogous situation where the twists are given on one of the L_i. Let $\tilde{Z}_k^{\langle n \rangle}$ be the subscheme of $\boldsymbol{Gr}(n,k) \times \boldsymbol{Gr}(n,k-1)$ whose S-valued points are the isomorphism classes of pairs $(\tilde{L}_1, \tilde{L}_2)$, where $\tilde{L}_1$ and $\tilde{L}_2$ are respectively subbundles of $\mathcal{O}_S^{\oplus n}$ of rank k and $k-1$ satisfying $\tilde{L}_2 \subseteq \tilde{L}_1$ and $\tilde{L}_2^{(p^2)} \subseteq \tilde{L}_1$. The relative Frobenius morphisms on the two Grassmannian factors induce two morphisms

$$\begin{array}{ccccc} Z_k^{\langle n \rangle} & \xrightarrow{\quad \varphi \quad} & \tilde{Z}_k^{\langle n \rangle} & \xrightarrow{\quad \hat{\varphi} \quad} & (Z_k^{\langle n \rangle})^{(p)} \\ (L_1, L_2) & \longmapsto & (L_1^{(p)}, L_2) & & \\ & & (\tilde{L}_1, \tilde{L}_2) & \longmapsto & (\tilde{L}_1, \tilde{L}_2^{(p)}), \end{array}$$

such that the composition is the relative Frobenius on $\tilde{Z}_k^{\langle n\rangle}$. Using a simple deformation computation, we see that φ has degree p^{n-k} and $\hat{\varphi}$ has degree p^{k-1}. Let $\tilde{\mathcal{L}}_1$ and $\tilde{\mathcal{L}}_2$ denote the universal subbundles on $\boldsymbol{Gr}(n,k)\times\boldsymbol{Gr}(n,k-1)$ when restricted to $\tilde{Z}_k^{\langle n\rangle}$; let $\tilde{\mathcal{Q}}_1$ and $\tilde{\mathcal{Q}}_2$ denote the universal quotients, respectively. We put

$$\tilde{\mathcal{E}}_k=(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^*\otimes\tilde{\mathcal{L}}_2^{(p^2)}\oplus(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*, \tag{5.4.1}$$

which is a vector bundle of rank $n-1$ on $\tilde{Z}_k^{\langle n\rangle}$.

Note that

$$\varphi^*(\tilde{\mathcal{E}}_k)=(\mathcal{L}_1^{(p)}/\mathcal{L}_2^{(p^2)})^*\otimes\mathcal{L}_2^{(p^2)}\oplus(\mathcal{L}_1^{(p)}/\mathcal{L}_2)\otimes\mathcal{Q}_1^{*,(p)}.$$

Comparing with $\mathcal{E}_k$, we see that $c_{n-1}(\varphi^*(\tilde{\mathcal{E}}_k))=p^{k-1}c_{n-1}(\mathcal{E}_k)$, where the factor p^{k-1} comes from the Frobenius twist on the first factor. Thus, we have

$$\begin{aligned}\int_{\tilde{Z}_k^{\langle n\rangle}}c_{n-1}(\tilde{\mathcal{E}}_k)&=(\deg\varphi)^{-1}\int_{Z_k^{\langle n\rangle}}c_{n-1}(\varphi^*(\tilde{\mathcal{E}}_k))\\&=p^{k-n}\int_{Z_k^{\langle n\rangle}}p^{k-1}c_{n-1}(\mathcal{E}_k)=p^{2k-n-1}N(n,k).\end{aligned} \tag{5.4.2}$$

Since $d(n-2\delta,k-\delta)+2k-n-1=2(k-\delta-1)(n-k-\delta+1)$, the theorem is in fact equivalent to the following (for each fixed k). $\square$

Proposition 5.5. *For* $1\le k\le(n+1)/2$, *we have*

$$\int_{\tilde{Z}_k^{\langle n\rangle}}c_{n-1}(\tilde{\mathcal{E}}_k)=(-1)^{n-1}\sum_{\delta=0}^{k-1}(n-2\delta)p^{2(k-\delta-1)(n-k-\delta+1)}\binom{n}{\delta}_{p^2}. \tag{5.5.1}$$

Remark 5.6. Before giving the proof of this proposition, we point out a variant of the construction of $\tilde{Z}_k^{\langle n\rangle}$. Let $\tilde{Z}_k^{\prime\langle n\rangle}$ be the subscheme of $\boldsymbol{Gr}(n,k)\times\boldsymbol{Gr}(n,k-1)$ whose S-valued points are the isomorphism classes of pairs $(\tilde{L}_1',\tilde{L}_2')$, where $\tilde{L}_1'$ and $\tilde{L}_2'$ are respectively subbundles of $\mathcal{O}_S^{\oplus n}$ of rank k and $k-1$ satisfying $\tilde{L}_2'\subseteq\tilde{L}_1'$ and $\tilde{L}_2'\subseteq\tilde{L}_1^{\prime(p^2)}$ (Note that the twist is on L_1' as opposed to be on L_2'). This is again a certain partial-Frobenius twist of $Z_k^{\langle n\rangle}$; it is smooth of dimension $n-1$. Define the universal subbundles and quotient bundles $\tilde{\mathcal{L}}_1'$, $\tilde{\mathcal{L}}_2'$, $\tilde{\mathcal{Q}}_1'$, and $\tilde{\mathcal{Q}}_2'$ similarly. We put

$$\tilde{\mathcal{E}}_k'=(\tilde{\mathcal{L}}_1'/\tilde{\mathcal{L}}_2')^*\otimes\tilde{\mathcal{L}}_2'\oplus(\tilde{\mathcal{L}}_1^{\prime(p^2)}/\tilde{\mathcal{L}}_2')\otimes(\tilde{\mathcal{Q}}_1^{\prime *})^{(p^2)}.$$

Using the same argument as above, we see that, for every fixed k,

$$\int_{\tilde{Z}_k^{\prime\langle n\rangle}}c_{n-1}(\tilde{\mathcal{E}}_k')=p^{n+1-2k}N(n,k).$$

Note that the exponent is different from (5.4.2). So Proposition 5.5 for each fixed k is equivalent to

$$\int_{\tilde{Z}_k'^{\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}_k') = (-1)^{n-1}\sum_{\delta=0}^{k-1}(n-2\delta)p^{2(k-\delta)(n-k-\delta)}\binom{n}{\delta}_{p^2},$$

as $2(k-\delta)(n-k-\delta) = d(n-2\delta, k-\delta)+n-2k+1$.

Proof of Proposition 5.5. We first prove it in the case of $k = 1, 2$ and then we explain an inductive process to deal with the general case.

When $k = 1$, $\tilde{Z}_1^{\langle n\rangle}$ classifies a line subbundle $\tilde{L}_1$ of $\mathcal{O}_S^{\oplus n}$ with no additional condition (as $\tilde{L}_2$ is zero); so $\tilde{Z}_1^{\langle n\rangle} \cong \mathbb{P}^{n-1}$ and $\tilde{\mathcal{L}}_1 = \mathcal{O}_{\mathbb{P}^{n-1}}(-1)$. The vector bundle $\tilde{\mathcal{E}}_1$ is equal to $\tilde{\mathcal{L}}_1 \otimes \tilde{\mathcal{Q}}_1^*$. It is straightforward to check that

$$c(\tilde{\mathcal{E}}_1) = \big(1+c_1(\mathcal{O}_{\mathbb{P}^{n-1}}(-1))\big)^n \quad \text{and hence} \quad \int_{\tilde{Z}_1^{\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}_1) = (-1)^{n-1}n;$$

the proposition is proved in this case.

When $k = 2$, we consider a forgetful morphism

$$\psi: \quad \tilde{Z}_2^{\langle n\rangle} \to \tilde{Z}_1^{\langle n\rangle}, \quad (\tilde{L}_1, \tilde{L}_2) \mapsto \tilde{L}_2.$$

This morphism is an isomorphism over the closed points $x \in \tilde{Z}_1^{\langle n\rangle}(\bar{\mathbb{F}}_p)$ for which $\tilde{L}_{2,x} \neq \tilde{L}_{2,x}^{(p^2)}$, because in this case $\tilde{L}_{1,x}$ is forced to be $\tilde{L}_{2,x} + \tilde{L}_{2,x}^{(p^2)}$. On the other hand, for a closed point $x \in \tilde{Z}_1^{\langle n\rangle}(\bar{\mathbb{F}}_p)$ where $\tilde{L}_{2,x} = \tilde{L}_{2,x}^{(p^2)}$, i.e., for $x \in \tilde{Z}_1^{\langle n\rangle}(\mathbb{F}_{p^2}) \cong \mathbb{P}^{n-1}(\mathbb{F}_{p^2})$, $\psi^{-1}(x)$ is the space classifying a line $\tilde{L}_1$ in $\bar{\mathbb{F}}_p^{\oplus n}/\tilde{L}_{2,x}$; so $\psi^{-1}(x) \simeq \mathbb{P}^{n-2}$. A simple tangent space computation shows that ψ is the blowup morphism of $\tilde{Z}_1^{\langle n\rangle} \cong \mathbb{P}^{n-1}$ at all of its $\mathbb{F}_{p^2}$-points. We use E to denote the exceptional divisors, which is a disjoint union of $\binom{n}{1}_{p^2}$ copies of $\mathbb{P}^{n-2}$.

Note that the vanishing of the morphism $\tilde{\mathcal{L}}_2 \to \tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)}$ defines the divisor E (as we can see using deformation); so

$$\mathcal{O}_{\tilde{Z}_2^{\langle n\rangle}}(E) \cong \tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)} \otimes \tilde{\mathcal{L}}_2^{-1}.$$

Put $\eta = c_1(\tilde{\mathcal{L}}_2) = \psi^* c_1(\mathcal{O}_{\mathbb{P}^{n-1}}(-1))$ and $\xi = c_1(E)$. Then

$$\begin{aligned} c(\tilde{\mathcal{E}}_2) &= c\big((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^* \otimes \tilde{\mathcal{L}}_2^{(p^2)}\big)\cdot c\big((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*\big) \\ &= (1-\xi+(p^2-1)\eta)\cdot(1+\xi+p^2\eta)^n/(1+\xi+(p^2-1)\eta), \end{aligned} \tag{5.6.1}$$

where the computation of the second term comes from the following two exact sequences

$$0 \to (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^* \to (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)^{\oplus n} \to (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{L}}_1^* \to 0;$$

$$0 \to \mathcal{O}_{\tilde{Z}_2^{\langle n\rangle}} \to (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{L}}_1^* \to (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{L}}_2^* \to 0.$$

Note that $\int_{\tilde{Z}_2^{\langle n\rangle}} \xi^i \eta^j = 0$ unless $(i, j) = (n-1, 0)$ or $(0, n-1)$, in which case

$$\int_{\tilde{Z}_2^{\langle n\rangle}} \eta^{n-1} = (-1)^{n-1} \quad \text{and} \quad \int_{\tilde{Z}_2^{\langle n\rangle}} \xi^{n-1} = (-1)^n \binom{n}{1}_{p^2}.$$

Here, to prove the last formula, we used the fact that the restriction of $\mathcal{O}_{\tilde{Z}_2^{\langle n\rangle}}(E)$ to each irreducible component $\mathbb{P}^{n-2}$ of E is isomorphic to $\mathcal{O}_{\mathbb{P}^{n-2}}(-1)$. So it suffices to compute

- the ξ^{n-1}-coefficient of (5.6.1), which is the same as the ξ^{n-1}-coefficient of $(1-\xi)(1+\xi)^{n-1}$ and is equal to $2-n$; and
- the η^{n-1}-coefficient of (5.6.1), which is the same as the η^{n-1}-coefficient of $(1+(p^2-1)\eta)(1+p^2\eta)^n/(1+(p^2-1)\eta) = (1+p^2\eta)^n$ and is equal to $np^{2(n-1)}$.

To sum up, we see that

$$\int_{\tilde{Z}_2^{\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}_2) = (-1)^{n-1} np^{2(n-1)} + (-1)^n (2-n) \binom{n}{1}_{p^2},$$

which is exactly (5.5.1) for $k = 2$.

In general, we make an induction on k. Assume that the proposition is proved for $k-1 \geq 1$ and we now prove the proposition for k (assuming that $k \leq \frac{1}{2}(n+1)$). By Remark 5.6, we get the similar intersection formula for $\tilde{\mathcal{E}}'_{k-1}$ on $\tilde{Z}'^{\langle n\rangle}_{k-1}$:

$$\int_{\tilde{Z}'^{\langle n\rangle}_{k-1}} c_{n-1}(\tilde{\mathcal{E}}'_{k-1}) = (-1)^{n-1} \sum_{\delta=0}^{k-2} (n-2\delta) p^{2(k-\delta-1)(n-k-\delta+1)} \binom{n}{\delta}_{p^2}. \tag{5.6.2}$$

We consider the moduli space W over $\mathbb{F}_{p^2}$ whose S-points are tuples $(\tilde{L}_1, \tilde{L}_2 = \tilde{L}'_2, \tilde{L}'_3)$, where $\tilde{L}_1$, $\tilde{L}_2$ and $\tilde{L}'_3$ are respectively subbundles of $\mathcal{O}_S^{\oplus n}$ of rank k, $k-1$ and $k-2$ satisfying $\tilde{L}'_3 \subset \tilde{L}_2 \subset \tilde{L}_1$ and $\tilde{L}'_3 \subset \tilde{L}_2^{(p^2)} \subset \tilde{L}_1$. It is easy to use deformation theory to check that W is a smooth variety of dimension $n-1$. There are two natural morphisms

$$\begin{tikzcd}[column sep=small] & W \arrow[dl, "\psi_{12}"'] \arrow[dr, "\psi_{23}"] & \\ \tilde{Z}_k^{\langle n\rangle} & & \tilde{Z}'^{\langle n\rangle}_{k-1} \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & (\tilde{L}_1, \tilde{L}_2 = \tilde{L}'_2, \tilde{L}'_3) \arrow[dl, mapsto] \arrow[dr, mapsto] & \\ (\tilde{L}_1, \tilde{L}_2) & & (\tilde{L}'_2, \tilde{L}'_3). \end{tikzcd}$$

Let E denote the subspace of W whose closed points $x \in W(\bar{\mathbb{F}}_p)$ are those such that $\tilde{L}_{2,x} = \tilde{L}_{2,x}^{(p^2)}$, i.e., $\tilde{L}_{2,x}$ is an $\mathbb{F}_{p^2}$-rational subspace of $\mathbb{F}_{p^2}^{\oplus n}$ of dimension $k-1$. It is clear that E is a disjoint union of $\binom{n}{k-1}_{p^2}$ copies (corresponding to the choices of $\tilde{L}_2$) of $\mathbb{P}^{n-k} \times \mathbb{P}^{k-2}$ (corresponding to the choice of $\tilde{L}_1$ and $\tilde{L}'_3$ respectively). It gives rise to a smooth divisor on W.

For a point $x \in (W \setminus E)(\bar{\mathbb{F}}_p)$, we have $\tilde{L}_{2,x} \neq \tilde{L}_{2,x}^{(p^2)}$ and hence it uniquely determines both $\tilde{L}_{1,x}$ and $\tilde{L}'_{3,x}$; so ψ_{12} and ψ_{23} are isomorphisms restricted to $W \setminus E$. On the other hand, when restricted to E, ψ_{12} contracts each copy of $\mathbb{P}^{n-k} \times \mathbb{P}^{k-2}$ of E into the first factor $\mathbb{P}^{n-k}$; whereas ψ_{23} contracts each copy of $\mathbb{P}^{n-k} \times \mathbb{P}^{k-2}$ of E into the second factor $\mathbb{P}^{k-2}$. It is clear from this (with a little bit of help from a deformation argument) that ψ_{12} is the blowup of $\tilde{Z}_k^{\langle n\rangle}$ along $\psi_{12}(E)$ and ψ_{23} is the blowup of $\tilde{Z}'^{\langle n\rangle}_{k-1}$ along $\psi_{23}(E)$; the divisor E is the exceptional divisor for both blowups.

A simple deformation theory argument shows that the normal bundle of E in W when restricted to each component $\mathbb{P}^{n-k} \times \mathbb{P}^{k-2}$ is $\mathcal{O}_{\mathbb{P}^{n-k}}(-1) \otimes \mathcal{O}_{\mathbb{P}^{k-2}}(-1)$. Moreover, we can characterize E as the zero locus of either one of the following natural homomorphisms

$$\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3 \to \tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2, \quad \tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3 \to \tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)}.$$

So as a line bundle over W, we have

$$\mathcal{O}_W(E) \cong (\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)^{-1} \otimes (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2) \cong (\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1} \otimes (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)}).$$

We want to compare

$$\begin{gathered}\int_{\tilde{Z}_k^{\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}_k) = \int_W c_{n-1}(\psi_{12}^*(\tilde{\mathcal{E}}_k)) \quad \text{and} \\ \int_{\tilde{Z}'^{\langle n\rangle}_{k-1}} c_{n-1}(\tilde{\mathcal{E}}'_{k-1}) = \int_W c_{n-1}(\psi_{23}^*(\tilde{\mathcal{E}}'_{k-1})).\end{gathered} \tag{5.6.3}$$

We will show that they differ by $(2k-n-2)(-1)^n\binom{n}{k-1}_{p^2}$ and this will conclude the proof of the proposition by inductive hypothesis (5.6.2). Indeed, we have

$$c\big(\psi_{12}^*(\tilde{\mathcal{E}}_k)\big) = c\big((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^* \otimes \tilde{\mathcal{L}}_2^{(p^2)}\big) \cdot c\big((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2) \otimes \tilde{\mathcal{Q}}_1^*\big), \tag{5.6.4}$$

$$c\big(\psi_{23}^*(\tilde{\mathcal{E}}'_{k-1})\big) = c\big((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^* \otimes \tilde{\mathcal{L}}'_3\big) \cdot c\big((\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3) \otimes \tilde{\mathcal{Q}}_2^{*,(p^2)}\big), \tag{5.6.5}$$

where $\tilde{\mathcal{Q}}_1$ and $\tilde{\mathcal{Q}}_2$ are the universal quotient vector bundles. Consider the following two exact sequences where the two last terms are identified:

$$\begin{tikzcd}[column sep=small]
 & & \mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^{-1}\otimes\tilde{\mathcal{L}}_2^{(p^2)} \arrow[d, leftrightarrow, "\cong"] & & \\
0 \arrow[r] & (\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}'_3 \arrow[r] & (\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}_2^{(p^2)} \arrow[r] & (\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes(\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3) \arrow[r] \arrow[d, leftrightarrow, "\cong"] & 0 \\
0 \arrow[r] & (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^* \arrow[r] & (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_2^{*,(p^2)} \arrow[r] \arrow[d, leftrightarrow, "\cong"] & (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes(\tilde{\mathcal{Q}}_2^{*,(p^2)}/\tilde{\mathcal{Q}}_1^*) \arrow[r] & 0. \\
 & & \mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)\otimes\tilde{\mathcal{Q}}_2^{*,(p^2)} & &
\end{tikzcd}$$

Here the right vertical isomorphism is given by

$$\begin{aligned}(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes(\tilde{\mathcal{Q}}_2^{*,(p^2)}/\tilde{\mathcal{Q}}_1^*) &\cong (\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^{-1}\\ &\cong \big((\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)\otimes\mathcal{O}_W(E)\big)\otimes\big((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)\otimes\mathcal{O}_W(E)\big)^{-1}\\ &\cong (\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes(\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3).\end{aligned}$$

From these two exact sequences we see that

$$\begin{aligned}&c\big((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}'_3\big)\cdot c\big(\mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)\otimes\tilde{\mathcal{Q}}_2^{*,(p^2)}\big)\\ &\qquad= c\big((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*\big)\cdot c\big(\mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^{-1}\otimes\tilde{\mathcal{L}}_2^{(p^2)}\big).\end{aligned}$$

Comparing this with (5.6.5) and (5.6.4), we get

$$\begin{aligned}&c_{n-1}(\psi_{12}^*(\tilde{\mathcal{E}}_k))-c_{n-1}(\psi_{23}^*(\tilde{\mathcal{E}}'_{k-1}))\\ &=\big(c_{k-1}((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^{-1}\otimes\tilde{\mathcal{L}}_2^{(p^2)})-c_{k-1}(\mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2^{(p^2)})^{-1}\otimes\tilde{\mathcal{L}}_2^{(p^2)})\big)\cdot c_{n-k}((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*)\\ &-c_{k-2}((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}'_3)\cdot\big(c_{n-k+1}((\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)\otimes\tilde{\mathcal{Q}}_2^{*,(p^2)})-c_{n-k+1}(\mathcal{O}_W(E)\otimes(\tilde{\mathcal{L}}_2^{(p^2)}/\tilde{\mathcal{L}}'_3)\otimes\tilde{\mathcal{Q}}_2^{*,(p^2)})\big).\end{aligned}$$

Recall that E is the exceptional divisor for the blow-up ψ_{12} centered at a disjoint union of $\mathbb{P}^{n-k}$; so $c_1(E)$ kills $\psi_{12}^*(A^i(\tilde{Z}_k^{\langle n\rangle}))$ for $i\geq n-k+1$. Similarly, $c_1(E)$ kills $\psi_{23}^*(A^i(\tilde{Z}_{k-1}^{\prime\langle n\rangle}))$ for $i\geq k-1$. As a result, we can rewrite the above complicated formula as

$$\begin{aligned}&c_{n-1}(\psi_{12}^*(\tilde{\mathcal{E}}_k))-c_{n-1}(\psi_{23}^*(\tilde{\mathcal{E}}'_{k-1}))\\ &\quad=-c_1(E)^{k-2}|_E\cdot c_{n-k}((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*)+c_{k-2}((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}'_3)\cdot c_1(E)^{n-k}|_E\\ &\quad=(-1)^{k-1}c_{n-k}((\tilde{\mathcal{L}}_1/\tilde{\mathcal{L}}_2)\otimes\tilde{\mathcal{Q}}_1^*)|_{\psi_{12}(E)}+(-1)^{n-k}c_{k-2}((\tilde{\mathcal{L}}_2/\tilde{\mathcal{L}}'_3)^{-1}\otimes\tilde{\mathcal{L}}'_3)|_{\psi_{23}(E)}.\end{aligned}$$

For the first term, over each $\mathbb{P}^{n-k}$ of $\psi_{12}(E)$, it is to take the top Chern class of the canonical subbundle of rank $n-k$ twisted by $\mathcal{O}_{\mathbb{P}^{n-k}}(-1)$; so the degree of the first term is $(-1)^{n-k}(n-k+1)$ on each $\mathbb{P}^{n-k}$. Similarly, for the second term, over each $\mathbb{P}^{k-2}$, it is the top Chern class of the canonical subbundle of rank $k-2$ twisted by $\mathcal{O}_{\mathbb{P}^{k-2}}(-1)$; so the degree of the second term is $(-1)^{k-2}(k-1)$ on each $\mathbb{P}^{k-2}$. To sum up, we have

$$\begin{aligned}&\int_W c_{n-1}(\psi_{12}^*(\tilde{\mathcal{E}}_k))-\int_W c_{n-1}(\psi_{23}^*(\tilde{\mathcal{E}}'_{k-1}))\\ &\qquad=(-1)^{k-1}(-1)^{n-k}(n-k+1)\binom{n}{k-1}_{p^2}+(-1)^{n-k}(-1)^{k-2}(k-1)\binom{n}{k-1}_{p^2}\\ &\qquad=(-1)^{n-1}(n-2k+2)\binom{n}{k-1}_{p^2}.\end{aligned}\tag{5.6.6}$$

So by the inductive hypothesis,

$$\begin{aligned}\int_{\tilde{Z}_k^{\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}_k) &\overset{(5.6.3)}{=} \int_W c_{n-1}(\psi_{12}^*(\tilde{\mathcal{E}}_k)) \\ &\overset{(5.6.6)}{=} \int_W c_{n-1}(\psi_{23}^*(\tilde{\mathcal{E}}_k)) + (-1)^{n-1}(n-2k+2)\binom{n}{k-1}_{p^2}. \\ &\overset{(5.6.3)}{=} \int_{\tilde{Z}_{k-1}^{\prime\langle n\rangle}} c_{n-1}(\tilde{\mathcal{E}}'_{k-1}) + (-1)^{n-1}(n-2k+2)\binom{n}{k-1}_{p^2} \\ &= (-1)^{n-1}\sum_{\delta=0}^{k-2}(n-2\delta)\,p^{2(k-\delta-1)(n-k-\delta+1)}\binom{n}{\delta}_{p^2} \\ &\qquad\qquad + (-1)^{n-1}(n-2k+2)\binom{n}{k-1}_{p^2} \\ &= (-1)^{n-1}\sum_{\delta=0}^{k-1}(n-2\delta)\,p^{2(k-\delta-1)(n-k-\delta+1)}\binom{n}{\delta}_{p^2}.\end{aligned}$$

This shows the statement of the proposition for k and hence concludes the proof. □

6. Intersection matrix of supersingular cycles on $\mathrm{Sh}_{1,n-1}$

Throughout this section, we fix an integer $n \geq 2$ and keep the notation as in Section 4. We will study the intersection theory of cycles Y_j for $1 \leq j \leq n$ on $\mathrm{Sh}_{1,n-1}$ considered in Section 4. For this, we may assume the following:

Hypothesis 6.1. We assume that the tame level structure K^p is taken sufficiently small so that Lemma 4.13 holds with $N = 2$.

6.2. ***Hecke correspondences on*** $\mathbf{Sh_{0,n}}$**.** Recall that we have an isomorphism

$$G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_\mathfrak{p}) \cong \mathbb{Q}_p^\times \times \mathrm{GL}_n(\mathbb{Q}_{p^2}).$$

Put $K_\mathfrak{p} = \mathrm{GL}_n(\mathcal{O}_{E_\mathfrak{p}})$ and $K_p = \mathbb{Z}_p^\times \times K_\mathfrak{p}$. The Hecke algebra $\mathbb{Z}[K_\mathfrak{p} \backslash \mathrm{GL}_n(E_\mathfrak{p})/K_\mathfrak{p}]$ can be viewed as a subalgebra of $\mathbb{Z}[K_p \backslash G(\mathbb{Q}_p)/K_p]$ (with trivial factor at the $\mathbb{Q}_p^\times$-component).

For $\gamma \in \mathrm{GL}_n(E_\mathfrak{p})$, the double coset $T_\mathfrak{p}(\gamma) := K_\mathfrak{p}\gamma K_\mathfrak{p}$ defines a Hecke correspondence on $\mathrm{Sh}_{0,n}$. It induces a set theoretic Hecke correspondence

$$T_\mathfrak{p}(\gamma) : \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p) \to \mathcal{S}(\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)),$$

where $\mathcal{S}(\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p))$ denotes the set of subsets of $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. By Remark 4.12, $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ is a union of $\#\ker^1(\mathbb{Q}, G_{0,n})$-isogeny classes of abelian varieties. Fix a base point $z_0 \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. Let

$$\Theta_{z_0} : \mathrm{Isog}(z_0) \xrightarrow{\sim} G_{0,n}(\mathbb{Q}) \backslash (G(\mathbb{A}^{\infty,p}) \times G(\mathbb{Q}_p))/K^p \times K_p.$$

be the bijection constructed as in Corollary 4.11. Write $K_{\mathfrak{p}}\gamma K_{\mathfrak{p}} = \coprod_{i\in I}\gamma_i K_{\mathfrak{p}}$. If $z \in \mathrm{Isog}(z_0)$ corresponds to the class of $(g^p, g_p) \in G(\mathbb{A}^{\infty,p}) \times G(\mathbb{Q}_p)$ with $g_p = (g_{p,0}, g_{\mathfrak{p}})$, then $T_{\mathfrak{p}}(\gamma)(z)$ consists of points in $\mathrm{Isog}(z_0)$ corresponding to the class of $(g^p, (g_{p,0}, g_{\mathfrak{p}}\gamma_i))$ for all $i \in I$.

Alternatively, $T_{\mathfrak{p}}(\gamma)$ has the following description. Write $z = (A, \lambda, \eta)$, and let $\mathbb{L}_z$ denote the $\mathbb{Z}_{p^2}$-free module $\tilde{\mathcal{D}}(A)_1^{\circ, F^2=p}$. Then a point $z' = (B, \lambda', \eta') \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ belongs to $T_{\mathfrak{p}}(\gamma)(z)$ if and only if there exists an $\mathcal{O}_D$-equivariant p-quasi-isogeny $\phi : B' \to B$ (i.e., $p^m\phi$ is an isogeny of p-power order for some integer m) such that

(1) $\phi^\vee \circ \lambda \circ \phi = \lambda'$,

(2) $\phi \circ \eta' = \eta$,

(3) $\phi_*(\mathbb{L}_{z'})$ is a lattice of $\mathbb{L}_z[1/p] = \mathbb{L}_z \otimes_{\mathbb{Z}_{p^2}} \mathbb{Q}_{p^2}$ with the property: there exists a $\mathbb{Z}_{p^2}$-basis $(e_1, \dots, e_n)$ for $\mathbb{L}_z$ such that $(e_1, \dots, e_n)\gamma$ is a $\mathbb{Z}_{p^2}$-basis for $\phi_*(\mathbb{L}_{z'})$.

When $\gamma = \mathrm{Diag}(p^{a_1}, \dots, p^{a_n})$ with $a_i \in \{-1, 0, 1\}$, For given z and z', such a ϕ is necessarily unique if it exists, by Lemma 4.13 (with $N = 2$). Therefore, $T_{\mathfrak{p}}(\gamma)(z)$ is in natural bijection with the set of $\mathbb{Z}_{p^2}$-lattices $\mathbb{L}' \subseteq \mathbb{L}_z[1/p]$ satisfying property (3) above.

For each integer i with $0 \le i \le n$, we put

$$T_{\mathfrak{p}}^{(i)} = T_{\mathfrak{p}}(\mathrm{Diag}(\underbrace{p, \dots, p}_{i}, \underbrace{1, \dots, 1}_{n-i})).$$

By the discussion above, one has a natural bijection

$$T_{\mathfrak{p}}^{(i)}(z) \xrightarrow{\sim} \{\mathbb{L}_{z'} \subseteq \mathbb{L}_z[1/p] \mid p\mathbb{L}_z \subseteq \mathbb{L}_{z'} \subseteq \mathbb{L}_z,\ \dim_{\mathbb{F}_{p^2}}(\mathbb{L}_z/\mathbb{L}_{z'}) = i\}$$

for $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$. Note that $T_{\mathfrak{p}}^{(0)} = \mathrm{id}$, and we put $S_{\mathfrak{p}} := T_{\mathfrak{p}}^{(n)}$. Then the Satake isomorphism implies $\mathbb{Z}[K_{\mathfrak{p}} \backslash \mathrm{GL}_n(E_{\mathfrak{p}})/K_{\mathfrak{p}}] \cong \mathbb{Z}[T_{\mathfrak{p}}^{(1)}, \dots, T_{\mathfrak{p}}^{(n-1)}, S_{\mathfrak{p}}, S_{\mathfrak{p}}^{-1}]$. More generally, for $0 \le a \le b \le n$, we put

$$R_{\mathfrak{p}}^{(a,b)} = T_{\mathfrak{p}}(\mathrm{Diag}(\underbrace{p^2, \dots, p^2}_{a}, \underbrace{p, \dots, p}_{b-a}, \underbrace{1, \dots, 1}_{n-b})).$$

Note that $R_{\mathfrak{p}}^{(0,i)} = T_{\mathfrak{p}}^{(i)}$, and $R_{\mathfrak{p}}^{(a,b)} S_{\mathfrak{p}}^{-1}$ is the Hecke operator

$$T_{\mathfrak{p}}(\mathrm{Diag}(\underbrace{p, \dots, p}_{a}, \underbrace{1, \dots, 1}_{b-a}, \underbrace{p^{-1}, \dots, p^{-1}}_{n-b})).$$

For the explicit relations between $R_{\mathfrak{p}}^{(a,b)}$ and $T_{\mathfrak{p}}^{(i)}$, see Proposition A.1.

6.3. ***Refined Gysin homomorphism.*** For an algebraic variety X over $\bar{\mathbb{F}}_p$ of pure dimension N and any integer $r \ge 0$, we write $A_r(X) = A^{N-r}(X)$ to denote the group of dimension r (codimension $N - r$) cycles in X modulo rational equivalence. Recall that the restriction of $\mathrm{pr}_j : Y_j \to \mathrm{Sh}_{1,n-1}$ to each $Y_{j,z}$ for $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ and

$1 \le j \le n$ is a regular closed immersion (into $\overline{\mathrm{Sh}}_{1,n-1}$). There is a well-defined Gysin homomorphism

$$\mathrm{pr}_j^! : A_{n-1}(\overline{\mathrm{Sh}}_{1,n-1}) \to A_0(\bar{Y}_j) = \bigoplus_{z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)} A_0(Y_{j,z}), \tag{6.3.1}$$

whose composition with the natural projection $A_0(\bar{Y}_j) \to A_0(Y_{j,z})$ is the refined Gysin map $(\mathrm{pr}_j|_{Y_{j,z}})^!$ defined in [Fulton 1998, 6.2] for regular immersions. Let $X \subseteq \overline{\mathrm{Sh}}_{1,n-1}$ be a closed subvariety of dimension $n-1$. Consider the Cartesian diagram

$$\begin{tikzcd} \bar{Y}_j \times_{\overline{\mathrm{Sh}}_{1,n-1}} X \arrow[r, "g_X"] \arrow[d, "g_j"'] & X \arrow[d] \\ \bar{Y}_j \arrow[r, "\mathrm{pr}_j"] & \overline{\mathrm{Sh}}_{1,n-1}. \end{tikzcd}$$

Assume that the restriction of g_X to each $Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X$ with $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ is a regular closed immersion as well. Then $\mathrm{pr}_j^!([X]) \in A_0(\bar{Y}_j)$ can be described as follows. Put $N_{Y_{j,z}}(\overline{\mathrm{Sh}}_{1,n-1}) := \mathrm{pr}_j^*(T_{\overline{\mathrm{Sh}}_{1,n-1}})/T_{Y_{j,z}}$, and we define $N_{Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X}(X)$ in a similar way. We define the *excess vector bundle* as

$$\mathcal{E}(Y_{j,z}, X) := g_j^* N_{Y_{j,z}}(\overline{\mathrm{Sh}}_{1,n-1}) / N_{Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X}(X).$$

This is a vector bundle on $Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X$. Let r be its rank function, which is equal to the dimension of $\bar{Y}_j \times_{\overline{\mathrm{Sh}}_{1,n-1}} X$ on each of its connected component. Then the excess intersection formula [Fulton 1998, 6.3] shows that

$$\mathrm{pr}_j^!([X]) = \sum_{z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)} \int_{Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X} c_r(\mathcal{E}(Y_{j,z}, X)), \tag{6.3.2}$$

where $c_r(\mathcal{E}(Y_{j,z}, X))$ is the top Chern class of $\mathcal{E}(Y_{j,z}, X)$ over $Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X$. The integration should be understood as the sum over all connected components of $Y_{j,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} X$ of the degrees of $c_r(\mathcal{E}(Y_{j,z}, X))$.

Proposition 6.4. *Let i, j be integers with $1 \le i \le j \le n$ and $z, z' \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$.*

(1) *The subvarieties $Y_{i,z}$ and $Y_{j,z'}$ of $\overline{\mathrm{Sh}}_{1,n-1}$ have nonempty intersection if and only if there exists an integer δ with $0 \le \delta \le \min\{n-j, i-1\}$ such that $z' \in R_{\mathfrak{p}}^{(j-i+\delta, n-\delta)} S_{\mathfrak{p}}^{-1}(z)$, or equivalently $z \in R_{\mathfrak{p}}^{(\delta, n+i-j-\delta)} S_{\mathfrak{p}}^{-1}(z')$, where $R_{\mathfrak{p}}^{(a,b)}$ and $S_{\mathfrak{p}}$ are the Hecke operators defined in Section 6.2.*

(2) *If the condition in* (1) *is satisfied for some δ, then $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is isomorphic to the variety $\bar{Z}_{i-\delta}^{\langle n+i-j-2\delta \rangle}$ defined in Section 5.2. Moreover, the excess vector bundles $\mathcal{E}(Y_{i,z}, Y_{j,z'})$ and $\mathcal{E}(Y_{j,z'}, Y_{i,z})$ are both isomorphic to the vector bundle* (5.2.1) *on $\bar{Z}_{i-\delta}^{\langle n+i-j-2\delta \rangle}$.*

Proof. Let $(\mathcal{B}_z, \lambda_z, \eta_z)$ and $(\mathcal{B}_{z'}, \lambda_{z'}, \eta_{z'})$ be the universal polarized abelian varieties on $\overline{\mathrm{Sh}}_{0,n}$ at z and z', respectively. Then $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is the moduli space of tuples $(A, \lambda, \eta, \phi, \phi')$ where $\phi : \mathcal{B}_z \to A$ and $\phi' : \mathcal{B}_{z'} \to A$ are isogenies such that $(A, \lambda, \eta, \mathcal{B}_z, \lambda_z, \eta_z, \phi)$ and $(A, \lambda, \eta, \mathcal{B}_{z'}, \eta_{z'}, \phi')$ are points of $Y_{i,z}$ and $Y_{j,z'}$ respectively.

Assume first that $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is nonempty, and let $(A, \lambda, \eta, \phi, \phi')$ be an $\bar{\mathbb{F}}_p$-valued point of it. Denote by $\tilde{\omega}^\circ_{A^\vee,k} \subseteq \tilde{\mathcal{D}}(A)^\circ_k$ for $k = 1, 2$ the inverse image of $\omega^\circ_{A^\vee/\bar{\mathbb{F}}_p,k} \subseteq H_1^{\mathrm{dR}}(A/\bar{\mathbb{F}}_p)^\circ_k \cong \tilde{\mathcal{D}}(A)^\circ_k / p\tilde{\mathcal{D}}(A)^\circ_k$. We identify $\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_k$ and $\tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_k$ with their images in $\tilde{\mathcal{D}}(A)^\circ_k$ via $\phi_{z,*,k}$ and $\phi_{z',*,k}$. Then we have a diagram of inclusions of $W(\bar{\mathbb{F}}_p)$-modules:

$$\begin{tikzcd}
 & & & \tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 \arrow[dr, hook, "\delta"] & & \\
p\tilde{\mathcal{D}}(A)^\circ_1 \arrow[r, hook, "1"] & \tilde{\omega}^\circ_{A^\vee,1} \arrow[r, hook, "n-j-\delta"] & \tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1 \arrow[ur, hook, "j-i+\delta"] \arrow[dr, hook, "\delta"'] & & \tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 + \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1 \arrow[r, hook, "i-\delta-1"] & \tilde{\mathcal{D}}(A)^\circ_1. \\
 & & & \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1 \arrow[ur, hook, "j-i+\delta"'] & &
\end{tikzcd} \tag{6.4.1}$$

Here the numbers on the arrows indicate the $\bar{\mathbb{F}}_p$-dimensions of the cokernel of the corresponding inclusions, which we shall compute below. By the definition of Y_i and Y_j, we have

$$\dim_{\bar{\mathbb{F}}_p}(\tilde{\mathcal{D}}(A)^\circ_1 / \tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1) = \dim_{\bar{\mathbb{F}}_p} \mathrm{Coker}(\phi_{*,1}) = i - 1,$$

and similarly, $\dim_{\bar{\mathbb{F}}_p}(\tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1 / \tilde{\omega}^\circ_{A^\vee,1}) = n - j$. Therefore, if we put

$$\delta = \dim_{\bar{\mathbb{F}}_p}\big(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 + \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1\big)/\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 = \dim_{\bar{\mathbb{F}}_p} \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1/\big(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1\big),$$

we have $0 \le \delta \le \min\{i-1, n-j\}$. Moreover, the quasi-isogeny $\phi_{z,z'} = \phi^{-1} \circ \phi' : \mathcal{B}_{z'} \to \mathcal{B}_z$ makes $\mathcal{B}_{z'}$ an element of $\mathrm{Isog}(z)$. We identify $\mathbb{L}_{z'}$ defined in (4.11.1) with a $\mathbb{Z}_{p^2}$-lattice of $\mathbb{L}_z[1/p]$ via $\phi_{z',z,*,1}$. Then

$$\dim_{\mathbb{F}_{p^2}}(\mathbb{L}_z \cap \mathbb{L}_{z'})/p\mathbb{L}_z = \dim_{\bar{\mathbb{F}}_p}(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1)/p\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1 = n + i - j - \delta.$$

Take a $\mathbb{Z}_{p^2}$-basis $(e_1, \dots, e_n)$ of $\mathbb{L}_z$ such that the image of $(e_{j-i+\delta+1}, \dots, e_n)$ in $\mathbb{L}_z/p\mathbb{L}_z$ form a basis of $(\mathbb{L}_z \cap \mathbb{L}_{z'})/p\mathbb{L}_z$ and such that $p^{-1}e_{n-\delta+1}, \dots, p^{-1}e_n$ form a basis of $(\mathbb{L}_z + \mathbb{L}_{z'})/\mathbb{L}_z$. Then

$$(pe_1, \dots, pe_{j-i+\delta}, e_{j-i+\delta+1}, \dots, e_{n-\delta}, p^{-1}e_{n-\delta+1}, \dots, p^{-1}e_n) \tag{6.4.2}$$

is a basis of $\mathbb{L}_{z'}$, that is $z' \in R_{\mathfrak{p}}^{(j-i+\delta, n-\delta)} S_{\mathfrak{p}}^{-1}(z)$ according to the convention of Section 6.2.

Conversely, assume that there exists δ with $1 \leq \delta \leq \min\{i-1, n-j\}$ such that the point $z' \in R_{\mathfrak{p}}^{(j-i+\delta,n-\delta)} S_{\mathfrak{p}}^{-1}(z)$. We have to prove statement (2), then the nonemptiness of $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ will follow automatically. Let $\phi_{z',z} : \mathcal{B}_{z'} \to \mathcal{B}_z$ be the unique quasi-isogeny which identifies $\mathbb{L}_{z'}$ with a $\mathbb{Z}_{p^2}$-lattice of $\mathbb{L}_z[1/p]$. By the definition of $R_{\mathfrak{p}}^{(j-i+\delta,n-\delta)} S_{\mathfrak{p}}^{-1}$, there exists a basis $e_1, \dots, e_n$ of $\mathbb{L}_z$ such that (6.4.2) is a basis of $\mathbb{L}_{z'}$. One checks easily that $p(\mathbb{L}_z + \mathbb{L}_{z'}) \subseteq \mathbb{L}_z \cap \mathbb{L}_{z'}$. We put

$$M_k = \big(\tilde{\mathcal{D}}(\mathcal{B}_z)_k^\circ \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})_k^\circ\big) / p\big(\tilde{\mathcal{D}}(\mathcal{B}_z)_k^\circ + \tilde{\mathcal{D}}(\mathcal{B}_{z'})_k^\circ\big)$$

for $k = 1, 2$. Then one has

$$\dim_{\bar{\mathbb{F}}_p}(M_k) = \dim_{\mathbb{F}_{p^2}}(\mathbb{L}_z \cap \mathbb{L}_{z'})/p(\mathbb{L}_z + \mathbb{L}_{z'}) = n + i - j - 2\delta.$$

The Frobenius and Verschiebung on $\tilde{\mathcal{D}}(\mathcal{B}_z)$ induce two bijective Frobenius semilinear maps $F : M_1 \to M_2$ and $V^{-1} : M_2 \to M_1$. We denote their linearizations by the same notation if no confusions arise. Let $Z_\delta(M_\bullet)$ be the moduli space which attaches to each locally noetherian $\bar{\mathbb{F}}_p$-scheme S the set of isomorphism classes of pairs (L_1, L_2), where $L_1 \subseteq M_1 \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S$ and $L_2 \subseteq M_2 \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S$ are subbundles of rank $i-\delta$ and $i-1-\delta$ respectively such that

$$L_2 \subseteq F(L_1^{(p)}), \quad V^{-1}(L_2^{(p)}) \subseteq L_1.$$

Note that there exists a basis $(\varepsilon_{k,1}, \dots, \varepsilon_{k,n+i-j-2\delta})$ of M_k for $k = 1, 2$ under which the matrices of F and V^{-1} are both identity. Indeed, by solving a system of equations of Artin–Schreier type, one can take a basis $(\varepsilon_{1,\ell})_{1 \leq \ell \leq n+i-j-2\delta}$ for M_1 such that

$$V^{-1}(F(\varepsilon_{1,\ell})) = \varepsilon_{1,\ell} \qquad \text{for all } 1 \leq \ell \leq n+i-j-2\delta.$$

We put $\varepsilon_{2,\ell} = F(\varepsilon_{1,\ell})$. Using these bases to identify both M_1 and M_2 with $\bar{\mathbb{F}}_p^{n+i-j-2\delta}$, it is clear that $Z_\delta(M_\bullet)$ is isomorphic to the variety $\overline{Z}_{i-\delta}^{\langle n+i-j-2\delta\rangle}$ considered in Section 5.2.

We have to establish an isomorphism between $Z_\delta(M_\bullet)$ and $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$. Let (L_1, L_2) be an S-point of $Z_\delta(M_\bullet)$. Note that there is a natural surjection

$$\big((\tilde{\mathcal{D}}(\mathcal{B}_z)_k^\circ \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})_k^\circ)/p\tilde{\mathcal{D}}(\mathcal{B}_z)_k^\circ\big) \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S \to M_k \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S.$$

We define $H_{z,k}$ for $k = 1, 2$ to be the inverse image of L_k under this surjection. Then $H_{z,k}$ can be naturally viewed as a subbundle of $\mathcal{D}(\mathcal{B}_z)_k^\circ \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S$ of rank $i+1-k$, and we have $H_{z,2} \subseteq F(H_{z,1}^{(p)})$ and $V^{-1}(H_{z,2}^{(p)}) \subseteq H_{z,1}$ since the pair (L_1, L_2) verifies similar properties. Therefore, $(L_1, L_2) \mapsto (\mathcal{B}_{z,S}, \lambda_{z,S}, \eta_{z,S}, H_{z,1}, H_{z,2})$ gives rise to a well-defined map $\varphi'_{i,z} : Z_\delta(M_\bullet) \to Y'_{i,z}$, where $(\mathcal{B}_{z,S}, \lambda_{z,S}, \eta_{z,S})$ is the base change of $(\mathcal{B}_z, \lambda_z, \eta_z)$ to S. Similarly, we have a morphism $\varphi'_{j,z'} : Z_\delta(M_\bullet) \to Y'_{j,z'}$ defined by $(L_1, L_2) \mapsto (\mathcal{B}_{z',S}, \lambda_{z',S}, \eta_{z',S}, H_{z',1}, H_{z',2})$, where $H_{z',k}$ is the inverse image of

L_k under the natural surjection:

$$\big((\tilde{\mathcal{D}}(\mathcal{B}_z)_k^\circ \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})_k^\circ)/p\tilde{\mathcal{D}}(\mathcal{B}_{z'})_k^\circ\big) \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S \to M_k \otimes_{\bar{\mathbb{F}}_p} \mathcal{O}_S.$$

By Proposition 4.8, we get two morphisms

$$\varphi_{i,z} : Z_\delta(M_\bullet) \to Y_{i,z}, \qquad \varphi_{j,z'} : Z_\delta(M_\bullet) \to Y_{j,z'}.$$

We claim that $\mathrm{pr}_i \circ \varphi_{i,z} = \mathrm{pr}_j \circ \varphi_{j,z}$, so that $(\varphi_{i,z}, \varphi_{j,z'})$ defines a map

$$\varphi : Z_\delta(M_\bullet) \to Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}.$$

Since $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is separated, the locus where $\mathrm{pr}_i \circ \varphi_{i,z}$ coincides with $\mathrm{pr}_j \circ \varphi_{j,z}$ is a closed subscheme of $Z_\delta(M_\bullet)$. As $Z_\delta(M_\bullet)$ is reduced, it is enough to show $\mathrm{pr}_i(\varphi_{i,z}(x)) = \mathrm{pr}_j(\varphi_{j,z}(x))$ for each closed geometric point $x = (L_1, L_2) \in Z_\delta(M_\bullet)(\bar{\mathbb{F}}_p)$. Let $(A, \lambda, \eta, \mathcal{B}_z, \lambda_z, \eta_z, \phi)$ and $(A', \lambda', \eta', \mathcal{B}_{z'}, \lambda_{z'}, \eta'_{z'}, \phi')$ be respectively the image of (L_1, L_2) under $\varphi_{i,z}$ and $\varphi_{j,z'}$. To prove the claim, we have to show that there is an isomorphism $(A, \lambda, \eta) \cong (A', \lambda', \eta')$ as objects of $\overline{\mathrm{Sh}}_{1,n-1}$. We identify $\tilde{\mathcal{D}}(\mathcal{B}_{z'})$, $\tilde{\mathcal{D}}(A)$, $\tilde{\mathcal{D}}(A')$ with $W(\bar{\mathbb{F}}_p)$-lattices of $\tilde{\mathcal{D}}(\mathcal{B}_z)[1/p]$ via the quasi-isogenies $\phi_{z',z} : \mathcal{B}_{z'} \to \mathcal{B}_z$, $\phi^{-1} : A \to \mathcal{B}_z$ and $\phi_{z',z}^{-1} \circ \phi' : A' \to \mathcal{B}_z$. Then by the construction of A (cf., the proof of Proposition 4.8), $\tilde{\mathcal{D}}(A)_1^\circ$ and $\tilde{\omega}^\circ_{A^\vee,1}$ fit into the diagram (6.4.1) such that there is a canonical isomorphism

$$\begin{aligned} L_1 &\cong \tilde{\omega}^\circ_{A^\vee,1}\big/p\big(\tilde{\mathcal{D}}(\mathcal{B}_z)_1^\circ + \tilde{\mathcal{D}}(\mathcal{B}_{z'})_1^\circ\big) \\ &\subseteq \big(\tilde{\mathcal{D}}(\mathcal{B}_z)_1^\circ \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})_1^\circ\big)\big/p\big(\tilde{\mathcal{D}}(\mathcal{B}_z)_1^\circ + \tilde{\mathcal{D}}(\mathcal{B}_{z'})_1^\circ\big) = M_1. \end{aligned} \tag{6.4.3}$$

Similarly, we have

$$\begin{aligned} L_2 &\cong p\tilde{\omega}^\circ_{A^\vee,2}/p\big(\tilde{\mathcal{D}}(\mathcal{B}_z)_2^\circ + \tilde{\mathcal{D}}(\mathcal{B}_{z'})_2^\circ\big) \\ &\subseteq \big(\tilde{\mathcal{D}}(\mathcal{B}_z)_2^\circ \cap \tilde{\mathcal{D}}(\mathcal{B}_{z'})_2^\circ\big)\big/p\big(\tilde{\mathcal{D}}(\mathcal{B}_z)_2^\circ + \tilde{\mathcal{D}}(\mathcal{B}_{z'})_2^\circ\big) = M_2. \end{aligned} \tag{6.4.4}$$

It is easy to see that such relations determine $\tilde{\mathcal{D}}(A)$ uniquely from (L_1, L_2). But the same argument shows that the same relations are satisfied with A replaced by A'. Hence, we see that the quasi-isogeny f induces an isomorphism between the Dieudonné modules of A and A'. As f is a p-quasi-isogeny, this implies immediately that f is an isomorphism of abelian varieties, proving the claim.

It remains to prove that $\varphi : Z_\delta(M_\bullet) \xrightarrow{\sim} Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is an isomorphism. It suffices to show that φ induces bijections on closed points and tangents spaces. The argument is similar to the proof of Proposition 4.8. Indeed, given a closed point $x = (A, \lambda, \eta, \phi, \phi')$ of $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$, one can construct a unique point $y = (L_1, L_2)$ of $Z_\delta(M_\bullet)$ with $\varphi(y) = x$ by the relations (6.4.3) and (6.4.4). It follows immediately that φ induces a bijection on closed points. Let x and y be as above. By the same argument as in Proposition 4.4, the tangent space of $Z_\delta(M_\bullet)$ at y is given by

$$T_{Z_\delta(M_\bullet),y} \cong (L_1/V^{-1}(L_2^{(p)}))^* \otimes (M_1/L_1) \oplus L_2^* \otimes F(L_1^{(p)})/L_2.$$

On the other hand, using Grothendieck–Messing deformation theory, one sees easily that the tangent space of $Y_{i,z}\times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ at x is given by

$$T_{Y_{i,z}\times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'},x}\cong \mathrm{Hom}_{\bar{\mathbb{F}}_p}\big(\omega^\circ_{A^\vee,1},\big(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_1\cap\tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1\big)\big/\tilde{\omega}^\circ_{A^\vee,1}\big)\\ \oplus\mathrm{Hom}_{\bar{\mathbb{F}}_p}\big(\tilde{\omega}^\circ_{A^\vee,2}\big/\big(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_2+\tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_2\big),\tilde{\mathcal{D}}(A)^\circ_2/\tilde{\omega}^\circ_{A^\vee,2}\big).$$

From (6.4.3) and (6.4.4), we see easily that

$$\omega^\circ_{A^\vee,1}\cong L_1/V^{-1}(L_2^{(p)}),\qquad \tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_1)/\tilde{\omega}^\circ_{A^\vee,1}\cong M_1/L_1,\\ \tilde{\omega}^\circ_{A^\vee,2}\big/\big(\tilde{\mathcal{D}}(\mathcal{B}_z)^\circ_2+\tilde{\mathcal{D}}(\mathcal{B}_{z'})^\circ_2\big)\cong L_2,\qquad \tilde{\mathcal{D}}(A)^\circ_2/\tilde{\omega}^\circ_{A^\vee,2}\cong F(L_1^{(p)})/L_2.$$

It follows that φ induces a bijection between $T_{Z_\delta(M_\bullet),y}$ and $T_{Y_{i,z}\times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'},x}$. This finishes the proof of Proposition 6.4. $\square$

6.5. ***Applications to cohomology.*** Recall that we have a morphism $\mathcal{JL}_j$ (4.16.1) for each $j=1,\dots,n$. We consider another map in the opposite direction:

$$\nu_j: H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1},\overline{\mathbb{Q}}_\ell(n-1))\xrightarrow{\mathrm{pr}_j^*} H^{2(n-1)}_{\mathrm{et}}(\bar{Y}_j,\overline{\mathbb{Q}}_\ell)\xrightarrow{\sim} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell),$$

where the second isomorphism is induced by the trace map

$$\mathrm{Tr}_{\mathrm{pr}'_j}: R^{2(n-1)}\,\mathrm{pr}'_{j,*}\,\overline{\mathbb{Q}}_\ell(n-1)\xrightarrow{\sim}\overline{\mathbb{Q}}_\ell.$$

For $1\le i,\ j\le n$, we define

$$m_{i,j}=\nu_j\circ\mathcal{JL}_i: H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell)\xrightarrow{\mathcal{JL}_i} H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1},\overline{\mathbb{Q}}_\ell(n-1))\xrightarrow{\nu_j} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell).$$

Putting all the morphisms $\mathcal{JL}_i$ and ν_j together, we get a sequence of morphisms:

$$\bigoplus_{i=1}^n H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell)\xrightarrow{\mathcal{JL}} H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1},\overline{\mathbb{Q}}_\ell(n-1))\\ \xrightarrow{\nu=(\nu_1,\dots,\nu_n)}\bigoplus_{j=1}^n H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell). \quad (6.5.1)$$

We see that the composed morphism above is given by the matrix $M=(m_{i,j})_{1\le i,j\le n}$, and we call it the *intersection matrix* of cycles Y_j on $\mathrm{Sh}_{1,n-1}$. All these morphisms are equivariant under the natural action of the Hecke algebra $\mathscr{H}(K^p,\overline{\mathbb{Q}}_\ell)$. We describe the intersection matrix in terms of the Hecke action of $\overline{\mathbb{Q}}_\ell[K_\mathfrak{p}\backslash\mathrm{GL}_n(E_\mathfrak{p})/K_\mathfrak{p}]$ on $H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell)$.

The group $H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n},\overline{\mathbb{Q}}_\ell)$ is the space of functions on $\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$ with values in $\overline{\mathbb{Q}}_\ell$. For $z\in\mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$, let e_z denote the characteristic function of z. Then the image of z under $K_\mathfrak{p}\gamma K_\mathfrak{p}$ for $\gamma\in\mathrm{GL}_n(E_\mathfrak{p})$ is

$$[K_\mathfrak{p}\gamma K_\mathfrak{p}]_*(e_z)=\sum_{z'\in T_\mathfrak{p}(\gamma)(z)} e_{z'},$$

where $T_{\mathfrak{p}}(\gamma)(z)$ means the set theoretic Hecke correspondence defined in Section 6.2. In the sequel, we will use the same notation $T_{\mathfrak{p}}(\gamma)$ to denote the action of $[K_{\mathfrak{p}}\gamma K_{\mathfrak{p}}]$ on $H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)$. In particular, we have Hecke operators $T^{(i)}_{\mathfrak{p}}$, $S_{\mathfrak{p}}$, $R^{(a,b)}_{\mathfrak{p}}$, ...

Proposition 6.6. *For $1 \le i \le j \le n$, we have*

$$m_{i,j} = \sum_{\delta=0}^{\min\{i-1,n-j\}} N(n+i-j-2\delta, i-\delta) R_{\mathfrak{p}}^{(j-i+\delta,n-\delta)} S_{\mathfrak{p}}^{-1},$$

$$m_{j,i} = \sum_{\delta=0}^{\min\{i-1,n-j\}} N(n+i-j-2\delta, i-\delta) R_{\mathfrak{p}}^{(\delta,n+i-j-\delta)} S_{\mathfrak{p}}^{-1},$$

where $N(n+i-j-2\delta, i-\delta)$ are the fundamental intersection numbers defined by (5.2.2).

Proof. We have a commutative diagram:

$$\begin{array}{ccccc} A_{n-1}(\bar{Y}_i) & \xrightarrow{\mathrm{pr}_{i,*}} & A_{n-1}(\overline{\mathrm{Sh}}_{1,n-1}) & \xrightarrow{\mathrm{pr}_j^!} & A_0(\bar{Y}_j) \\ \downarrow{\scriptstyle \mathrm{cl}} & & \downarrow{\scriptstyle \mathrm{cl}} & & \downarrow{\scriptstyle \mathrm{cl}} \\ H^0_{\mathrm{et}}(\bar{Y}_i, \overline{\mathbb{Q}}_\ell) & \xrightarrow{\mathrm{Gys}_{\mathrm{pr}_i}} & H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1)) & \xrightarrow{\mathrm{pr}_j^*} & H^{2(n-1)}_{\mathrm{et}}(\bar{Y}_j, \overline{\mathbb{Q}}_\ell). \end{array} \tag{6.6.1}$$

Here, the vertical arrows are cycle class maps, and $\mathrm{pr}_j^!$ is the refined Gysin map defined in (6.3.1). For $z \in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)$, the image of e_z under $m_{i,j}$ is given by

$$\begin{aligned} m_{i,j}(e_z) &= \mathrm{Tr}_{\mathrm{pr}'_j} \mathrm{pr}_j^* \mathrm{Gys}_{\mathrm{pr}_i} \mathrm{cl}([Y_{i,z}]) = \mathrm{Tr}_{\mathrm{pr}'_j}\big(\mathrm{cl}(\mathrm{pr}_j^! \mathrm{pr}_{i,*}[Y_{i,z}])\big) \\ &= \mathrm{Tr}_{\mathrm{pr}'_j}\Bigg(\sum_{z'\in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)} \mathrm{cl}\big(c_{r(z,z')}(\mathcal{E}(Y_{j,z'}, Y_{i,z}))\big)\cdot \mathrm{cl}(Y_{j,z'} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{i,z})\Bigg) \\ &= \sum_{z'\in \mathrm{Sh}_{0,n}(\bar{\mathbb{F}}_p)} \Bigg(\int_{Y_{j,z'}\times_{\mathrm{Sh}_{1,n-1}} Y_{i,z}} c_{r(z,z')}(\mathcal{E}(Y_{j,z'}, Y_{i,z}))\Bigg) e_{z'}, \end{aligned}$$

where $r(z,z')$ is the rank of $\mathcal{E}(Y_{j,z'}, Y_{i,z})$, and we used (6.3.2) in the second step. Indeed, Proposition 6.4(1) says that the schematic intersection $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'}$ is smooth, so the closed immersion $Y_{i,z} \times_{\overline{\mathrm{Sh}}_{1,n-1}} Y_{j,z'} \hookrightarrow Y_{j,z'}$ is a regular immersion and the assumptions for (6.3.2) are thus satisfied here.

By Proposition 6.4(1), $e_{z'}$ has a nonzero contribution to the summation above if and only if there exists an integer δ with $0 \le \delta \le \min\{i-1, n-j\}$ such that $z' \in R_{\mathfrak{p}}^{(j-i+\delta,n-\delta)} S_{\mathfrak{p}}^{-1}(z)$. In that case, Proposition 6.4(2) implies that the coefficient of $e_{z'}$ is nothing but the fundamental intersection number $N(n+i-j-2\delta, i-\delta)$ defined in (5.2.2). The formula for $m_{i,j}$ now follows immediately. The formula for $m_{j,i}$ is proved in the same manner. $\square$

If we express $m_{i,j}$ in terms of the elementary Hecke operators $T_{\mathfrak{p}}^{(k)}$, we get the following.

Theorem 6.7. *Put* $d(n,k)=(2k-1)n-2k(k-1)-1$ *for integers* $1\le k\le n$. *Then, for* $1\le i\le j\le n$, *we have*

$$m_{i,j}=\sum_{\delta=0}^{\min\{i-1,n-j\}}(-1)^{n+1+i-j}(n+i-j-2\delta)\,p^{d(n+i-j-2\delta,i-\delta)}T_{\mathfrak{p}}^{(j-i+\delta)}T_{\mathfrak{p}}^{(n-\delta)}S_{\mathfrak{p}}^{-1},$$

$$m_{j,i}=\sum_{\delta=0}^{\min\{i-1,n-j\}}(-1)^{n+1+i-j}(n+i-j-2\delta)\,p^{d(n+i-j-2\delta,i-\delta)}T_{\mathfrak{p}}^{(\delta)}T_{\mathfrak{p}}^{(n+i-j-\delta)}S_{\mathfrak{p}}^{-1}.$$

Proof. We prove only the statement for $m_{i,j}$, and that for $m_{j,i}$ is similar. By Proposition A.1 in Appendix A, the right hand side of the first formula above is

$$\sum_{\delta=0}^{\min\{i-1,n-j\}}(-1)^{n+1-i-j}(n+i-j-2\delta)\,p^{d(n+i-j-2\delta,i-\delta)}\cdot\Bigl(\sum_{k=0}^{\delta}\binom{n+i-j-2\delta+2k}{k}_{p^2}R_{\mathfrak{p}}^{(j-i+\delta-k,n-\delta+k)}S_{\mathfrak{p}}^{-1}\Bigr)=\sum_{r=0}^{\min\{i-1,n-j\}}(\star)R_{\mathfrak{p}}^{(j-i+r,n-r)}S_{\mathfrak{p}}^{-1}.$$

Here, we have put $r=\delta-k$, and the expression $\star$ in the parentheses is

$$\star=\sum_{k=0}^{\min\{i-1-r,n-j-r\}}(-1)^{n+1+i-j}(n+i-j-2r-2k)\cdot p^{d(n+i-j-2r-2k,i-r-k)}\binom{n+i-j-2r}{k}_{p^2}=N(n+i-j-2r,i-r).$$

Here, the last equality is Theorem 5.3. The statement for $m_{i,j}$ now follows from Proposition 6.6. ☐

Example 6.8. We write down explicitly the intersection matrices when n is small.

(1) Consider first the case $n=2$. This case is essentially the same as the Hilbert quadratic case studied in [Tian and Xiao 2014], and the intersection matrix can be written:

$$M=\begin{pmatrix}-2p & T_{\mathfrak{p}}^{(1)}\\ T_{\mathfrak{p}}^{(1)}S_{\mathfrak{p}}^{-1} & -2p\end{pmatrix}.$$

(2) When $n = 3$, Theorem 6.7 gives

$$M = \begin{pmatrix} 3p^2 & -2pT_{\mathfrak{p}}^{(1)} & T_{\mathfrak{p}}^{(2)} \\ -2pT_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & 3p^4 + T_{\mathfrak{p}}^{(1)}T_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & -2pT_{\mathfrak{p}}^{(1)} \\ T_{\mathfrak{p}}^{(1)}S_{\mathfrak{p}}^{-1} & -2pT_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & 3p^2 \end{pmatrix}.$$

(3) The intersection matrix for $n = 4$ can be written:

$$M = \begin{pmatrix} -4p^3 & 3p^2T_{\mathfrak{p}}^{(1)} & -2pT_{\mathfrak{p}}^{(2)} & T_{\mathfrak{p}}^{(3)} \\ 3p^2T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1} & -4p^7-2pT_{\mathfrak{p}}^{(1)}T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1} & 3p^4T_{\mathfrak{p}}^{(1)}+T_{\mathfrak{p}}^{(2)}T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1} & -2pT_{\mathfrak{p}}^{(2)} \\ -2pT_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & 3p^4T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1}+T_{\mathfrak{p}}^{(1)}T_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & -4p^7-2pT_{\mathfrak{p}}^{(1)}T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1} & 3p^2T_{\mathfrak{p}}^{(1)} \\ T_{\mathfrak{p}}^{(1)}S_{\mathfrak{p}}^{-1} & -2pT_{\mathfrak{p}}^{(2)}S_{\mathfrak{p}}^{-1} & 3p^2T_{\mathfrak{p}}^{(3)}S_{\mathfrak{p}}^{-1} & -4p^3 \end{pmatrix}.$$

6.9. ***Proof of Theorem 4.18(1).*** Let $\pi \in \mathscr{A}_K$ as in the statement of Theorem 4.18(1). Consider the $(\pi^p)^{K^p}$-isotypic direct factor of the $\mathscr{H}(K^p, \overline{\mathbb{Q}}_\ell)$-equivariant sequence (6.5.1):

$$\bigoplus_{i=1}^{n} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_{\pi^p} \xrightarrow{\mathcal{JL}_\pi} H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))_{\pi^p} \xrightarrow{\nu_\pi} \bigoplus_{j=1}^{n} H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_{\pi^p}. \quad (6.9.1)$$

In particular, when $i = j = 1$, $\nu_1 \circ \mathcal{JL}_1$ is given by multiplication by $-np^{n-1}$. So the π^p-isotypic component of (6.9.1) is nonzero. This implies that π^p appears in $H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))$, i.e., there exist admissible irreducible representations π'_p of $G_{1,n-1}(\mathbb{Q}_p)$ and π'_∞ of $G_{1,n-1}(\mathbb{R})$, which is cohomological in degree $n-1$, such that $\pi^p \otimes \pi'_p \otimes \pi'_\infty$ is a cuspidal automorphic representation $\pi' \otimes \pi'_\infty$ of $G_{1,n-1}(\mathbb{A}_\mathbb{Q})$. By Lemma 4.17, $\pi' \simeq \pi$ satisfies Hypothesis 2.5(2) for $a_\bullet = (1, n-1)$. Thus, taking the π^p-isotypic component of (6.9.1) is the same as taking its π-isotypic component. From now on, we use subscript π in places of subscript π^p.

If $a_{\mathfrak{p}}^{(i)}$ denotes the eigenvalues of $T_{\mathfrak{p}}^{(i)}$ on $\pi_{\mathfrak{p}}^{K_{\mathfrak{p}}}$ for each $1 \le i \le n$, then $T_{\mathfrak{p}}^{(i)}$ acts as the scalar $a_{\mathfrak{p}}^{(i)}$ on all the terms in (6.9.1). Therefore, $\nu_\pi \circ \mathcal{JL}_\pi$ is given by the matrix M_π, which is obtained by replacing $T_{\mathfrak{p}}^{(i)}$ by $a_{\mathfrak{p}}^{(i)}$ in each entry of M. By definition, the $\alpha_{\pi_{\mathfrak{p}},i}$ are the roots of the Hecke polynomial (2.6.2):

$$X^n + \sum_{i=1}^{n} (-1)^i p^{i(i-1)} a_{\mathfrak{p}}^{(i)} X^{n-i}.$$

Then Theorem 4.18(1) follows easily from the following.

Lemma 6.10. *We have*

$$\det(M_\pi) = \pm p^{\frac{n(n^2-1)}{3}} \frac{\prod_{i<j}(\alpha_{\pi_{\mathfrak{p}},i} - \alpha_{\pi_{\mathfrak{p}},j})^2}{\left(\prod_{i=1}^{n} \alpha_{\pi_{\mathfrak{p}},i}\right)^{n-1}}.$$

Here, $\pm$ means that the formula holds up to sign. In particular, $\nu_\pi \circ \mathcal{JL}_\pi$ is an isomorphism if the $\alpha_{\pi_{\mathfrak{p}},i}$ are distinct.

Proof. Put $\beta_i = \alpha_{\pi_{\mathfrak{p}},i}/p^{n-1}$ for $1 \le i \le n$. For $i = 1, \dots, n$, let s_i be the i-th elementary symmetric polynomial in $\beta_1, \dots, \beta_n$. Then we have $a_{\mathfrak{p}}^{(i)} = p^{i(n-i)} s_i$. It follows from Theorem 6.7 that the (i, j)-entry of M_π with $1 \le i \le j \le n$ is given by

$$m_{i,j}(\pi) = s_n^{-1} \sum_{\delta=0}^{\min\{i-1,n-j\}} (-1)^{n+1+i-j}(n+i-j-2\delta) \\ \cdot p^{d(n+i-j-2\delta,i-\delta)+(j-i+\delta)(n+i-j-\delta)+\delta(n-\delta)} s_{j-i+\delta} s_{n-\delta}.$$

A direct computation shows that the exponent index on p in each term above is independent of δ, and is equal to $e(i, j) := (n+1)(i+j-1)-(i^2+j^2)$. The same holds when $i > j$. In summary, we get $m_{i,j}(\pi) = s_n^{-1} p^{e(i,j)} m'_{i,j}(\pi)$ with

$$m'_{i,j}(\pi) = \begin{cases} \sum_{\delta=0}^{\min\{i-1,n-j\}} (-1)^{n+1+i-j}(n+i-j-2\delta) s_{j-i+\delta} s_{n-\delta}, & \text{if } i \le j, \\ \sum_{\delta=0}^{\min\{j-1,n-i\}} (-1)^{n+1+j-i}(n+j-i-2\delta) s_\delta s_{n+j-i-\delta}, & \text{if } i > j. \end{cases}$$

For any n-permutation σ, we have

$$\sum_{i=1}^{n} e(i, \sigma(i)) = \frac{n(n^2-1)}{3}.$$

Thus we get $\det(M_\pi) = p^{n(n^2-1)/3} s_n^{-n} \det(m'_{i,j}(\pi))$. The rest of the computation is purely combinatorial, which is the case $q = -1$ of Theorem B.1 in Appendix B. □

Remark 6.11. We point out that the determinant of the intersection matrix computed by Theorem B.1 holds with an auxiliary variable q. A similar phenomenon also appeared in the case of Hilbert modular varieties [Tian and Xiao 2014], where the computation was related to the combinatorial model of periodic semimeanders. These motivate us to ask, out of curiosity, whether there might be some quantum version of the construction of cycles, or even Conjecture 2.12, possibly for the geometric Langlands setup.

6.12. ***Proof of Theorem 4.18(2).*** Given Theorem 4.18(1), it suffices to prove that

$$n \dim H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_\pi \ge \dim H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell(n-1))^{\mathrm{fin}}_\pi. \tag{6.12.1}$$

Actually, by (2.4.1) and (2.6.3), we have

$$H^0_{\mathrm{et}}(\overline{\mathrm{Sh}}_{0,n}, \overline{\mathbb{Q}}_\ell)_\pi = \pi^K \otimes R_{(0,n),\ell}(\pi), \quad H^{2(n-1)}_{\mathrm{et}}(\overline{\mathrm{Sh}}_{1,n-1}, \overline{\mathbb{Q}}_\ell)_\pi = \pi^K \otimes R_{(1,n-1),\ell}(\pi).$$

Write $\pi_p = \pi_{p,0} \otimes \pi_{\mathfrak{p}}$ as a representation of $G(\mathbb{Q}_p) \simeq \mathbb{Q}_p^\times \times \mathrm{GL}_n(E_{\mathfrak{p}})$. Let $\chi_{\pi_{p,0}}$: $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_{p^2}) \to \overline{\mathbb{Q}}_\ell^\times$ denote the character sending Frob_{p^2} to $\pi_{p,0}(p^2)$, and let $\rho_{\pi_{\mathfrak{p}}}$

be as in (2.6.1). According to (2.6.3), up to semisimplification, we have

$$[R_{(0,n),\ell}(\pi)] = \#\ker^1(\mathbb{Q}, G_{0,n}) m_{0,n}(\pi) \big[\wedge^n \rho_{\pi_{\mathfrak{p}}} \otimes \chi_{\pi_{p,0}}^{-1} \otimes \overline{\mathbb{Q}}_\ell\big(\tfrac{1}{2}n(n-1)\big)\big], \tag{6.12.2}$$

$$[R_{(1,n-1),\ell}(\pi)] = \#\ker^1(\mathbb{Q}, G_{1,n-1}) m_{0,n}(\pi) \big[\rho_{\pi_{\mathfrak{p}}} \otimes \wedge^{n-1} \rho_{\pi_{\mathfrak{p}}} \otimes \chi_{\pi_{p,0}}^{-1} \otimes \overline{\mathbb{Q}}_\ell\big(\tfrac{1}{2}(n-1)(n-2)\big)\big]. \tag{6.12.3}$$

Note that

$$\dim\Big(\rho_{\pi_{\mathfrak{p}}} \otimes \wedge^{n-1} \rho_{\pi_{\mathfrak{p}}} \otimes \chi_{\pi_{p,0}}^{-1} \otimes \overline{\mathbb{Q}}_\ell\Big(\frac{(n-1)(n-2)}{2}\Big)\Big)^{\mathrm{fin}} = \sum_{\zeta} \dim(\rho_{\pi_{\mathfrak{p}}} \otimes \wedge^{n-1} \rho_{\pi_{\mathfrak{p}}})^{\mathrm{Frob}_{p^2} = p^{n(n-1)}\zeta},$$

where the superscript "fin" means taking the subspace on which $\mathrm{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_{p^2})$ acts through a finite quotient, and ζ runs through all roots of unity. If $\alpha_{\pi_{\mathfrak{p}},i}/\alpha_{\pi_{\mathfrak{p}},j}$ is not a root of unity for any pair $i \neq j$, the right hand side above is equal to the sum of the multiplicities of $\prod_{i=1}^n \alpha_{\pi_{\mathfrak{p}},i} = p^{n(n-1)}\zeta$ as eigenvalues of $(\rho_{\pi_{\mathfrak{p}}} \otimes \wedge^{n-1} \rho_{\pi_{\mathfrak{p}}})(\mathrm{Frob}_{p^2})$, which is n. Therefore, under these conditions on the $\alpha_{\pi_{\mathfrak{p}},i}$, we have by (6.12.3)

$$\dim R_{(1,n-1),\ell}(\pi)^{\mathrm{fin}} \leq n \cdot \#\ker^1(\mathbb{Q}, G_{1,n-1}) \cdot m_{1,n-1}(\pi),$$

and the equality holds if Frob_{p^2} is semisimple on $R_{(1,n-1),\ell}(\pi)$. On the other hand, we have from (6.12.2)

$$\dim R_{(0,n),\ell}(\pi) = \#\ker^1(\mathbb{Q}, G_{0,n}) \cdot m_{0,n}(\pi).$$

By a result of White [2012, Theorem E], the multiplicity $m_{a_\bullet}(\pi)$ above is equal to 1 for $a_\bullet = (1, n-1)$ and $a_\bullet = (0, n)$. Now the inequality (6.12.1) follows immediately from this and the fact that $\#\ker^1(\mathbb{Q}, G_{1,n-1}) = \#\ker^1(\mathbb{Q}, G_{0,n})$. This finishes the proof of Theorem 4.18(2). □

7. Construction of cycles in the case of $G(U(r, s) \times U(s, r))$

We keep the notation of Section 3.6. In this section, we will give the construction of certain cycles on Shimura varieties for $G(U(r, s) \times U(s, r))$. We always assume that $s \geq r$.

7.1. ***Description of the cycles in terms of Dieudonné modules.*** Let δ be a nonnegative integer with $\delta \leq r$. We consider the case of Conjecture 2.12 when $n = r + s$, $a_1 = r$, $a_2 = s$, $b_1 = r - \delta$, and $b_2 = s + \delta$. The representation $r_{a_\bullet}$ of GL_n involved is

$$r_{a_\bullet} = \wedge^r \mathrm{Std} \otimes \wedge^s \mathrm{Std}\,.$$

The weight λ of Conjecture 2.12 is

$$\lambda = (\underbrace{2, \dots, 2}_{r-\delta}, \underbrace{1, \dots, 1}_{s-r+2\delta}, \underbrace{0, \dots, 0}_{r-\delta}).$$

By elementary calculation of representations of GL_n, the multiplicity of λ in $r_{a_\bullet}$ is $m_\lambda(a_\bullet) = \binom{s-r+2\delta}{\delta}$. Then Conjecture 2.12 thus predicts the existence of $\binom{s-r+2\delta}{\delta}$ cycles Y_j on $\mathrm{Sh}_{r,s}$, each of dimension

$$\tfrac{1}{2}(\dim \mathrm{Sh}_{r,s} + \dim \mathrm{Sh}_{r-\delta,s+\delta}) = \tfrac{1}{2}(2rs + 2(r-\delta)(s+\delta)) = 2rs - (s-r)\delta - \delta^2,$$

and each admits a rational map to $\mathrm{Sh}_{r-\delta,s+\delta}$. The principal goal of this section is to construct these cycles, at least conjecturally. We start with the description in terms of the Dieudonné modules at closed points.

Consider the interval $[r-\delta, s+\delta]$; it contains $s-r+2\delta$ unit segments with integer endpoints. We will parametrize the cycles on the Shimura variety by the subsets of these $s-r+2\delta$ unit segments of cardinality δ. There are exactly $\binom{s-r+2\delta}{\delta}$ such subsets. Let $\boldsymbol{j}$ be one of them. Then we can write the union of all the segments in $\boldsymbol{j}$ as

$$[j_{1,1}, j_{1,2}] \cup [j_{2,1}, j_{2,2}] \cup \cdots \cup [j_{\epsilon,1}, j_{\epsilon,2}] \tag{7.1.1}$$

such that all $j_{\alpha,i}$ are integers,

$$r-\delta \le j_{1,1} < j_{1,2} < j_{2,1} < j_{2,2} < \cdots < j_{\epsilon,1} < j_{\epsilon,2} \le s+\delta,$$

and we have $\sum_{\alpha=1}^{\epsilon}(j_{\alpha,2} - j_{\alpha,1}) = \delta$. For notational convenience, we put $j_{0,1} = j_{0,2} = 0$.

We define $Z_{\boldsymbol{j}}$ to be the subset of $\bar{\mathbb{F}}_p$-points z of $\mathrm{Sh}_{r,s}$ such that the reduced Dieudonné modules $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1$ and $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2$ contain submodules $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$ satisfying (3.2.1) for $m = \epsilon$, i.e.,

$$p^\epsilon \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i \subseteq \tilde{\mathcal{E}}_i, \quad F(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{3-i}, \quad \text{and} \quad V(\tilde{\mathcal{E}}_i) \subseteq \tilde{\mathcal{E}}_{3-i}, \quad \text{for } i = 1, 2,$$

and the following condition for $i = 1, 2$:

$$\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i / \tilde{\mathcal{E}}_i \simeq (W(\bar{\mathbb{F}}_p)/p^\epsilon)^{\oplus j_{1,i}} \oplus (W(\bar{\mathbb{F}}_p)/p^{\epsilon-1})^{\oplus (j_{2,i}-j_{1,i})} \oplus \cdots \\ \cdots \oplus (W(\bar{\mathbb{F}}_p)/p)^{\oplus (j_{\epsilon,i}-j_{\epsilon-1,i})}. \tag{7.1.2}$$

We refer to the toy model discussed in Example 7.3 for the motivation of this condition. For technical reasons, we will not prove the set $Z_{\boldsymbol{j}}$ is the set of $\bar{\mathbb{F}}_p$-points of a closed subscheme of $\mathrm{Sh}_{r,s}$; instead we prove that a closely related subset of $Z_{\boldsymbol{j}}$ is. See Remark 7.5.

Applying Proposition 3.2 with $m = \delta$, the submodules $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$ give rise to a polarized abelian variety $(\mathcal{A}'_z, \lambda'_z)$ over z with an $\mathcal{O}_D$-action and an $\mathcal{O}_D$-equivariant isogeny $\mathcal{A}'_z \to \mathcal{A}_z$. Moreover, by (3.2.2), we have

$$\begin{aligned}\dim \omega^\circ_{\mathcal{A}'^\vee_z/\bar{\mathbb{F}}_p,1} &= \dim \omega^\circ_{\mathcal{A}^\vee_z/\bar{\mathbb{F}}_p,1} + \sum_{\alpha=0}^{\epsilon-1} \big((\epsilon-\alpha)(j_{\alpha+1,1}-j_{\alpha,1})-(\epsilon-\alpha)(j_{\alpha+1,2}-j_{\alpha,2})\big)\\ &= r-\delta\end{aligned}$$

and similarly $\dim \omega^\circ_{\mathcal{A}'^\vee_z/\bar{\mathbb{F}}_p,2} = s+\delta$. So $\mathcal{A}'_z$ satisfies the moduli problem for $\mathrm{Sh}_{r-\delta,s+\delta}$; this suggests a geometric relationship between $Z_{\boldsymbol{j}}$ and $\mathrm{Sh}_{r-\delta,s+\delta}$ that we make precise in Definition 7.4.

We make an immediate remark that when $\delta = r$, the abelian variety $\mathcal{A}_z$ coming from a point z of $Z_{\boldsymbol{j}}$ is isogenous to an abelian variety $\mathcal{A}'_z$ that is a moduli object for the Shimura variety $\mathrm{Sh}_{0,n}$. Thus both $\mathcal{A}'_z$ and $\mathcal{A}_z$ are supersingular. So every $Z_{\boldsymbol{j}}$ is contained in the supersingular locus of $\mathrm{Sh}_{r,s}$. In fact, we shall show in Theorem 7.8 that the supersingular locus of $\mathrm{Sh}_{r,s}$ is exactly the union of these $Z_{\boldsymbol{j}}$.

7.2. ***Towards a moduli interpretation.*** We need to reinterpret in a more geometric manner the Dieudonné-theoretic condition defining $Z_{\boldsymbol{j}}$. For $\alpha = 0, \dots, \epsilon$, we define submodules

$$\tilde{\mathcal{E}}_{\alpha,1} := \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_1 \quad\text{and}\quad \tilde{\mathcal{E}}_{\alpha,2} := \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2 \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_2$$

of $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1$ and $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2$. They are easily seen to satisfy condition (3.2.1) with $m=\alpha$. Thus, Proposition 3.2 generates a polarized abelian variety $(A_\alpha, \lambda_\alpha)$ with $\mathcal{O}_D$-action and an $\mathcal{O}_D$-equivariant isogeny $A_\alpha \to \mathcal{A}_z$, where

$$\begin{aligned} r_\alpha &:= \dim \omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,1} = r - \sum_{\alpha'=1}^{\alpha}(j_{\alpha',2}-j_{\alpha',1}) \quad\text{and}\\ s_\alpha &:= \dim \omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,2} = n - \dim \omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,1}\end{aligned} \tag{7.2.1}$$

by the formula (3.2.2). In particular $r_0 = r,\ s_0 = s,\ r_\epsilon = r-\delta$ and $s_\epsilon = s+\delta$.

In fact, applying Proposition 3.2 (with $m=1$) to the sequence of inclusions

$$\tilde{\mathcal{E}}_i = \tilde{\mathcal{E}}_{\epsilon,i} \subset \tilde{\mathcal{E}}_{\epsilon-1,i} \subset \cdots \subset \tilde{\mathcal{E}}_{0,i} = \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i,$$

we obtain a sequence of isogenies (each with p-torsion kernels):

$$\mathcal{A}'_z = A_\epsilon \xrightarrow{\phi_\epsilon} A_{\epsilon-1} \xrightarrow{\phi_{\epsilon-1}} \cdots \xrightarrow{\phi_1} A_0 = \mathcal{A}_z. \tag{7.2.2}$$

We have $\ker\phi_\alpha \subseteq A_\alpha[p]$, so that there exists a unique isogeny $\psi_\alpha : A_{\alpha-1} \to A_\alpha$ such that $\psi_\alpha\phi_\alpha = p\cdot \mathrm{id}_{A_\alpha}$ and $\phi_\alpha\psi_\alpha = p\cdot\mathrm{id}_{A_{\alpha-1}}$.

For each α, the cokernel of the induced map on cohomology

$$\phi_{\alpha,*,i} : H_1^{\mathrm{dR}}(A_\alpha/\bar{\mathbb{F}}_p)^\circ_i \to H_1^{\mathrm{dR}}(A_{\alpha-1}/\bar{\mathbb{F}}_p)^\circ_i$$
$$(\text{resp. } \psi_{\alpha,*,i} : H_1^{\mathrm{dR}}(A_{\alpha-1}/\bar{\mathbb{F}}_p)^\circ_i \to H_1^{\mathrm{dR}}(A_\alpha/\bar{\mathbb{F}}_p)^\circ_i)$$

is canonically isomorphic to $\tilde{\mathcal{E}}_{\alpha-1,i}/\tilde{\mathcal{E}}_{\alpha,i}$ (resp. $\tilde{\mathcal{E}}_{\alpha,i}/p\tilde{\mathcal{E}}_{\alpha-1,i}$), which has dimension $j_{\alpha,i}$ (resp. $n-j_{\alpha,i}$) over $\bar{\mathbb{F}}_p$ by a straightforward computation using (7.1.2).

The upshot is that all these numeric information of the chain of isogenies (7.2.2) can be used to reconstruct $\tilde{\mathcal{E}}_i$ inside $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i$. This idea will be made precise after this important example.

Example 7.3. We give a good toy model for the isogenies of Dieudonné modules. This is the inspiration of the construction of this section. We start with the Dieudonné module $\tilde{\mathcal{D}}(A_\epsilon)^\circ_1 = \bigoplus_{i=1}^n W(\bar{\mathbb{F}}_p)\boldsymbol{e}_j$ and $\tilde{\mathcal{D}}(A_\epsilon)^\circ_2 = \bigoplus_{j=1}^n W(\bar{\mathbb{F}}_p)\boldsymbol{f}_j$. The maps $V_1 : \tilde{\mathcal{D}}(A_\epsilon)^\circ_1 \to \tilde{\mathcal{D}}(A_\epsilon)^\circ_2$ and $V_2 : \tilde{\mathcal{D}}(A_\epsilon)^\circ_2 \to \tilde{\mathcal{D}}(A_\epsilon)^\circ_1$, with respect to the given bases, are given by the diagonal matrices

$$\mathrm{Diag}(\underbrace{1,\dots,1}_{s+\delta},\underbrace{p,\dots,p}_{r-\delta}) \quad \text{and} \quad \mathrm{Diag}(\underbrace{1,\dots,1}_{r-\delta},\underbrace{p,\dots,p}_{s+\delta}),$$

respectively. Using the isogenies ϕ_α we may naturally identify $\tilde{\mathcal{D}}(A_\alpha)^\circ_i$ as lattices in $\tilde{\mathcal{D}}(A_\epsilon)^\circ_i[1/p]$ with induced Frobenius and Verschiebung morphisms. For our toy model, we choose

$$\begin{aligned}\tilde{\mathcal{D}}(A_\alpha)^\circ_1 = \mathrm{Span}_{W(\bar{\mathbb{F}}_p)}\Big\{&\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_1,\dots,\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_{j_{\alpha+1,1}},\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{e}_{j_{\alpha+1,1}+1},\dots,\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{e}_{j_{\alpha+2,1}},\\ &\frac{1}{p^{\epsilon-\alpha-2}}\boldsymbol{e}_{j_{\alpha+2,1}+1},\dots,\frac{1}{p}\boldsymbol{e}_{j_{\epsilon,1}},\boldsymbol{e}_{j_{\epsilon,1}+1},\dots,\boldsymbol{e}_n\Big\};\\ \tilde{\mathcal{D}}(A_\alpha)^\circ_2 = \mathrm{Span}_{W(\bar{\mathbb{F}}_p)}\Big\{&\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{f}_1,\dots,\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{f}_{j_{\alpha+1,2}},\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{f}_{j_{\alpha+1,2}+1},\dots,\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{f}_{j_{\alpha+2,2}},\\ &\frac{1}{p^{\epsilon-\alpha-2}}\boldsymbol{f}_{j_{\alpha+2,2}+1},\dots,\frac{1}{p}\boldsymbol{f}_{j_{\epsilon,2}},\boldsymbol{f}_{j_{\epsilon,2}+1},\dots,\boldsymbol{f}_n\Big\}.\end{aligned}$$

In particular, the Verschiebung $V_1 : \tilde{\mathcal{D}}(A_\alpha)^\circ_1 \to \tilde{\mathcal{D}}(A_\alpha)^\circ_2$ with respect to the bases above is given by

$$\mathrm{Diag}(\underbrace{1,\dots,1}_{j_{\alpha+1,1}},***\cdots***,\underbrace{p,\dots,p}_{r-\delta}),$$

where the $***$ part is p if the place is in $[j_{\alpha',1}+1,\, j_{\alpha',2}]$ for some $\alpha' \geq \alpha$, and is 1 otherwise. Similarly, the Verschiebung $V_2 : \tilde{\mathcal{D}}(A_0)^\circ_2 \to \tilde{\mathcal{D}}(A_0)^\circ_1$ with respect to the bases above is given by

$$\mathrm{Diag}(\underbrace{1,\dots,1}_{r-\delta},\underbrace{p,\dots,p}_{j_{\alpha+1,1}-r+\delta},***\cdots***,\underbrace{p,\dots,p}_{n-j_{\alpha_{\epsilon,2}}}),$$

where the $***$ part is 1 if the place is in $[j_{\alpha',1}+1,\, j_{\alpha',2}]$ for some $\alpha' \geq \alpha$, and is p otherwise.

So the sheaf of differentials is given by

$$\omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,1} = \mathrm{Span}_{\bar{\mathbb{F}}_p}\Big\{\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_1,\dots,\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_{r-\delta},\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_{j_{\alpha,1}+1},\dots,\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{e}_{j_{\alpha,2}},$$
$$\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{e}_{j_{\alpha+1,1}+1},\dots,\frac{1}{p}\boldsymbol{e}_{j_{\epsilon-1,2}},\boldsymbol{e}_{j_{\epsilon,1}+1},\dots,\boldsymbol{e}_{j_{\epsilon,2}}\Big\};$$
$$\omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,2} = \mathrm{Span}_{\bar{\mathbb{F}}_p}\Big\{\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{f}_1,\dots,\frac{1}{p^{\epsilon-\alpha}}\boldsymbol{f}_{j_{\alpha+1,1}},\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{f}_{j_{\alpha+1,2}+1},\dots,\frac{1}{p^{\epsilon-\alpha-1}}\boldsymbol{f}_{j_{\alpha+2,1}},$$
$$\frac{1}{p^{\epsilon-\alpha-2}}\boldsymbol{f}_{j_{\alpha+2,1}},\dots,\frac{1}{p}\boldsymbol{f}_{j_{\epsilon,1}},\boldsymbol{f}_{j_{\epsilon,2}+1},\dots,\boldsymbol{f}_{s+\delta-1}\Big\}.$$

Definition 7.4. Let $\boldsymbol{j}$ be as above. Define the numbers $j_{\alpha,i}$ as in (7.1.1) and the numbers r_α, s_α as in (7.2.1). Let $\underline{Y}_{\boldsymbol{j}}$ be the functor taking a locally noetherian $\mathbb{F}_{p^2}$-scheme S to the set of isomorphism classes of tuples

$$(A_0,\dots,A_\epsilon,\lambda_0,\dots,\lambda_\epsilon,\eta_0,\dots,\eta_\epsilon,\phi_1,\dots,\phi_\epsilon,\psi_1,\dots,\psi_\epsilon) \tag{7.4.1}$$

such that:

(1) for each α, $(A_\alpha,\lambda_\alpha,\eta_\alpha)$ is an S-point of $\mathrm{Sh}_{r_\alpha,s_\alpha}$;

(2) for each α, ϕ_α is an $\mathcal{O}_D$-isogeny $A_\alpha\to A_{\alpha-1}$, with kernel contained in $A_\alpha[p]$, which is compatible with the polarizations in the sense that $p\lambda_\alpha=\phi_\alpha^\vee\circ\lambda_{\alpha-1}\circ\phi_\alpha$ and with the tame level structures in the sense that $\phi_\alpha\circ\eta_\alpha=\eta_{\alpha-1}$;

(3) ψ_α is the isogeny $A_{\alpha-1}\to A_\alpha$ such that $\phi_\alpha\psi_\alpha=p\cdot\mathrm{id}_{A_\alpha}$ and $\psi_\alpha\phi_\alpha=p\cdot\mathrm{id}_{A_{\alpha-1}}$;

(4) the cokernel of the induced map $\phi^{\mathrm{dR}}_{\alpha,*,i}:H_1^{\mathrm{dR}}(A_\alpha/S)_i^\circ\to H_1^{\mathrm{dR}}(A_{\alpha-1}/S)_i^\circ$ is a locally free $\mathcal{O}_S$-module of rank $j_{\alpha,i}$ for each α and $i=1,2$;

(5) the cokernel of the induced map $\psi^{\mathrm{dR}}_{\alpha,*,i}:H_1^{\mathrm{dR}}(A_{\alpha-1}/S)_i^\circ\to H_1^{\mathrm{dR}}(A_\alpha/S)_i^\circ$ is a locally free $\mathcal{O}_S$-module of rank $n-j_{\alpha,i}$ for each α and $i=1,2$;[18]

(6) for each α, $\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,2})$ is contained in $\omega^\circ_{A^\vee_\alpha/S,2}$;

(7) for each α, the $(r_{\alpha-1}-r_\alpha+r_\epsilon+1)$-st Fitting ideal of the cokernel of $\phi^{\mathrm{dR}}_{\alpha,*,1}:\omega^\circ_{A^\vee_\alpha/S,1}\to\omega^\circ_{A^\vee_{\alpha-1}/S,1}$ is zero, or equivalently, Zariski locally on S, if we represent the map $\phi^{\mathrm{dR}}_{\alpha,*,1}:\omega^\circ_{A^\vee_\alpha/S,1}\to\omega^\circ_{A^\vee_{\alpha-1}/S,1}$ by an $r_{\alpha-1}\times r_\alpha$-matrix (after choosing local bases), then all $(r_\alpha-r_\epsilon+1)\times(r_\alpha-r_\epsilon+1)$-minors vanish.

(8) the $(r_\alpha-r_\epsilon+1)$-st Fitting ideal of the cokernel of $\psi^{\mathrm{dR}}_{\alpha,*,1}:\omega^\circ_{A^\vee_{\alpha-1}/S,1}\to\omega^\circ_{A^\vee_\alpha/S,1}$ is zero for each α.

Note that conditions (6)–(8) are all closed conditions. So the moduli problem $\underline{Y}_{\boldsymbol{j}}$ is represented by a proper scheme $Y_{\boldsymbol{j}}$ of finite type over $\mathbb{F}_{p^2}$. The moduli space $Y_{\boldsymbol{j}}$ admits natural maps to $\mathrm{Sh}_{r,s}$ and $\mathrm{Sh}_{r-\delta,s+\delta}$ by sending the tuple (7.4.1) to

[18] This is in fact a corollary of (2) and (4).

(A_0, λ_0, η_0) and $(A_\epsilon, \lambda_\epsilon, \eta_\epsilon)$, respectively.

$$\begin{tikzcd} & Y_{\boldsymbol{j}} \arrow[dl, "\mathrm{pr}_{\boldsymbol{j}}"'] \arrow[dr, "\mathrm{pr}'_{\boldsymbol{j}}"] & \\ \mathrm{Sh}_{r,s} & & \mathrm{Sh}_{r-\delta,s+\delta} \end{tikzcd}$$

We also point out that conditions (2) and (3) together imply that, for each α and $i = 1, 2$, we have $\mathrm{Im}(\psi^{\mathrm{dR}}_{\alpha,*,i}) = \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,i})$ and $\mathrm{Im}(\phi^{\mathrm{dR}}_{\alpha,*,i}) = \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha,*,i})$. We shall freely use this property later.

Remark 7.5. Conditions (6)–(8) in Definition 7.4 are satisfied by the toy model in Example 7.3. They did not appear in moduli problem in Section 4.2 because they trivially hold in that case. The purpose of keeping these conditions in the moduli problem and carefully formulating them is so that the moduli space may hope to have the correct irreducible components. We think the picture is the following: $Z_{\boldsymbol{j}}$ is probably or at least heuristically the set of $\bar{\mathbb{F}}_p$-points of a closed subscheme of $\mathrm{Sh}_{r,s}$. But this scheme has many irreducible components, which may have overlaps with other $Z_{\boldsymbol{j}'}$. Conditions (6)–(8) will help select one irreducible component that is "special" for $\boldsymbol{j}$. When taking the union of all images of the $Y_{\boldsymbol{j}}$, we should still get the union of the $Z_{\boldsymbol{j}}$. This is verified in the case of supersingular locus (i.e., $r = \delta$) in Theorem 7.8.

Notation 7.6. Let $Y_{\boldsymbol{j}}$ as above. It will be convenient to introduce some dummy notation:

- ϕ_0 is the identity map on A_0;
- ψ_ϵ is the identity map on A_ϵ.

We use $Y^\circ_{\boldsymbol{j}}$ to denote the open subscheme of $Y_{\boldsymbol{j}}$ representing the functor that takes a locally noetherian $\mathbb{F}_{p^2}$-scheme S to the subset of isomorphism classes of tuples

$$(A_0, \dots, A_\epsilon, \lambda_0, \dots, \lambda_\epsilon, \eta_0, \dots, \eta_\epsilon, \phi_1, \dots, \phi_\epsilon, \psi_1, \dots, \psi_\epsilon)$$

of $Y_{\boldsymbol{j}}(S)$ such that

(i) for each $\alpha = 1, \dots, \epsilon$, the sum $\phi_{\alpha,*,2}(\omega^\circ_{A^\vee_\alpha/S,2}) + \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha-1,*,2})$ is an $\mathcal{O}_S$-subbundle of $H^{\mathrm{dR}}_1(A_{\alpha-1}/S)^\circ_2$ of rank

$$\operatorname{rank}\omega_{A^\vee_\alpha/S,2} - \operatorname{rank}\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,2}) + \operatorname{rank}\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha-1,*,2}) = s_\alpha - j_{\alpha,2} + j_{\alpha-1,2},$$

(ii) for each $\alpha = 1, \dots, \epsilon$, $\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1}) + \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1})$ is an $\mathcal{O}_S$-subbundle of rank

$$\operatorname{rank}\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1}) + \operatorname{rank}\mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1}) = j_{\alpha,1} + (n - j_{\alpha+1,1}),$$

(iii) for each α, the cokernel of $\phi^{\mathrm{dR}}_{\alpha,*,1} : \omega^\circ_{A^\vee_\alpha/S,1} \to \omega^\circ_{A^\vee_{\alpha-1}/S,1}$ is a locally free $\mathcal{O}_S$-module of rank $r_{\alpha-1} - (r_\alpha - r_\epsilon)$,

(iv) for each α, the cokernel of $\psi^{\mathrm{dR}}_{\alpha,*,1} : \omega^\circ_{A^\vee_{\alpha-1}/S,1} \to \omega^\circ_{A^\vee_\alpha/S,1}$ is a locally free $\mathcal{O}_S$-module of rank $r_\alpha - r_\epsilon$.

We note that the ranks in conditions (i) and (ii) are maximal possible and the ranks in conditions (iii) and (iv) are minimal possible, under the conditions in Definition 7.4. So Y_j° is an open subscheme of Y_j.

We point out an additional benefit of having conditions (ii)–(iv). By (iii), $\omega^\circ_{A_\alpha^\vee/S,1} \cap \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1})$ is an $\mathcal{O}_S$-subbundle of $\omega^\circ_{A_\alpha^\vee/S,1}$ of rank r_ϵ, for $\alpha = 1, \dots, \epsilon$; by (iv), $\omega^\circ_{A_\alpha^\vee/S,1} \cap \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1})$ is an $\mathcal{O}_S$-subbundle of $\omega^\circ_{A_\alpha^\vee/S,1}$ of rank $r_\alpha - r_\epsilon$, for $\alpha = 0, \dots, \epsilon - 1$. Combining these two rank estimates and condition (ii) which implies that $\mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1})$ and $\mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1})$ are disjoint subbundles, we arrive at a *direct sum* decomposition

$$\omega^\circ_{A_\alpha^\vee/S,1} = (\omega^\circ_{A_\alpha^\vee/S,1} \cap \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1})) \oplus (\omega^\circ_{A_\alpha^\vee/S,1} \cap \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1})), \tag{7.6.1}$$

for $\alpha = 1, \dots, \epsilon - 1$; and we know that $\omega^\circ_{A_0^\vee/S,1} \cap \mathrm{Ker}(\psi^{\mathrm{dR}}_{1,*,1})$ has rank $r_0 - r_\epsilon = \delta$ and $\omega^\circ_{A_\epsilon^\vee/S,1} \subseteq \mathrm{Ker}(\phi^{\mathrm{dR}}_{\epsilon,*,1})$.

We shall show below in Theorem 7.7 that Y_j° is smooth. Unfortunately, we do not know how to prove the nonemptiness of Y_j°, nor do we know if some Y_j is completely contained in some other Y_j; but the fact that the Dieudonné modules in Example 7.3 satisfy conditions (i)–(iv) above is good evidence for this nonemptiness. Of course, if one can compute the intersection matrix in the sense of Theorem 6.7 and calculate the determinant, one can then probably show that these Y_j are essentially different. But the difficulties of this computation lie in understanding the singularities at $Y_j \setminus Y_j^\circ$, which seems to be very combinatorially involved.

Theorem 7.7. *Each Y_j° is smooth of dimension $rs + (r-\delta)(s+\delta)$ (if not empty).*

Proof. Let $\hat{R}$ be a noetherian $\mathbb{F}_{p^2}$-algebra and $\hat{I} \subset \hat{R}$ an ideal such that $\hat{I}^2 = 0$. Put $R = \hat{R}/\hat{I}$. Say we want to lift an R-point

$$(A_0, \dots, A_\epsilon, \lambda_0, \dots, \lambda_\epsilon, \eta_0, \dots, \eta_\epsilon, \phi_1, \dots, \phi_\epsilon, \psi_1, \dots, \psi_\epsilon)$$

of Y_j° an $\hat{R}$-point and we try to compute the corresponding tangent space. By Serre–Tate and Grothendieck–Messing deformation theory we recalled in Theorem 3.4, it is enough to lift, for $i = 1, 2$ and each $\alpha = 0, \dots, \epsilon$, the differentials $\omega^\circ_{A_\alpha^\vee/R,i} \subseteq H_1^{\mathrm{dR}}(A_\alpha/R)_i^\circ$ to a subbundle $\hat{\omega}_{\alpha,i} \subseteq H_1^{\mathrm{cris}}(A_\alpha/\hat{R})_i^\circ$ such that

(a) $\phi^{\mathrm{cris}}_{\alpha,*,i}(\hat{\omega}_{\alpha,i}) \subseteq \hat{\omega}_{\alpha-1,i}$ and $\psi^{\mathrm{cris}}_{\alpha,*,i}(\hat{\omega}_{\alpha-1,i}) \subseteq \hat{\omega}_{\alpha,i}$ (so that both ϕ_α and ψ_α are lifted, which would automatically imply $\mathrm{Ker}(\phi_\alpha) \in A_\alpha[p]$),

(b) $\hat{\omega}_{\alpha,2} \supseteq \mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,2})$, and

(c) the $\hat{R}$-modules $\hat{\omega}_{\alpha-1,1}/\phi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}_{\alpha,1})$ and $\hat{\omega}_{\alpha,1}/\psi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}_{\alpha-1,1})$ are flat and of rank $r_{\alpha-1} - (r_\alpha - r_\epsilon)$ and $r_\alpha - r_\epsilon$, respectively.

We shall see that condition (i) of Notation 7.6 is automatic. Also, condition (ii) already holds: since $H_1^{\mathrm{cris}}(A_\alpha/\hat{R})_1^\circ/(\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,1}) + \mathrm{Ker}(\psi^{\mathrm{cris}}_{\alpha+1,*,1}))$ is locally generated by $j_{\alpha+1,1} - j_{\alpha,1}$ elements after modulo $\hat{I}$, it is so prior to modulo $\hat{I}$

by Nakayama's lemma. Note that rank of $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,1})$ and $\mathrm{Ker}(\psi^{\mathrm{cris}}_{\alpha+1,*,1})$ and the number of the generators of the quotient above add up to exactly n; it follows that $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,1}) + \mathrm{Ker}(\psi^{\mathrm{cris}}_{\alpha+1,*,1})$ is a direct sum and the sum is a subbundle of $H_1^{\mathrm{cris}}(A_\alpha/\hat{R})^\circ_1$.

We separate the discussion of lifts at q_1 and q_2, and show that the tangent space $T_{Y_j^\circ}$ is isomorphic to $T_1 \oplus T_2$ for the contributions T_1 and T_2 from the two places. We first look at q_2, as it is easier. Note that condition (b) $\hat{\omega}_{\alpha,2} \supseteq \mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,2}) = \mathrm{Im}(\psi^{\mathrm{cris}}_{\alpha,*,2})$ automatically implies that $\psi^{\mathrm{cris}}_{\alpha,*,2}(\hat{\omega}_{\alpha-1,2}) \subseteq \hat{\omega}_{\alpha,2}$; so we can proceed as follows:

Step 0: First lift $\omega^\circ_{A^\vee_\epsilon/R,2}$ to a subbundle $\hat{\omega}_{\epsilon,2}$ of $H_1^{\mathrm{cris}}(A_\epsilon/\hat{R})^\circ_2$ so that it contains $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\epsilon,*,2})$,

Step 1: then lift $\omega^\circ_{A^\vee_{\epsilon-1}/R,2}$ to a subbundle $\hat{\omega}_{\epsilon-1,2}$ of $H_1^{\mathrm{cris}}(A_{\epsilon-1}/\hat{R})^\circ_2$ so that it contains $\phi^{\mathrm{cris}}_{\epsilon,*,2}(\hat{\omega}_{\epsilon,2}) + \mathrm{Ker}(\phi^{\mathrm{cris}}_{\epsilon-1,*,2})$,

Step(s) $\boldsymbol{\alpha}$: then lift $\omega^\circ_{A^\vee_{\epsilon-\alpha}/R,2}$ to a subbundle $\hat{\omega}_{\epsilon-\alpha,2}$ of $H_1^{\mathrm{cris}}(A_{\epsilon-\alpha}/\hat{R})^\circ_2$ so that it contains $\phi^{\mathrm{cris}}_{\epsilon-\alpha+1,*,2}(\hat{\omega}_{\epsilon-\alpha+1,2}) + \mathrm{Ker}(\phi^{\mathrm{cris}}_{\epsilon-\alpha,*,2})$,

Step $\boldsymbol{\epsilon}$: finally lift $\omega^\circ_{A^\vee_0/R,2}$ to a subbundle $\hat{\omega}_{0,2}$ of $H_1^{\mathrm{cris}}(A_0/\hat{R})^\circ_2$ so that it contains $\phi^{\mathrm{cris}}_{1,*,2}(\hat{\omega}_{1,2})$.

At Step 0, the choices form a torsor for the group

$$\mathrm{Hom}_R(\omega^\circ_{A^\vee_\epsilon/R,2}/\,\mathrm{Ker}(\phi^{\mathrm{dR}}_{\epsilon,*,2}), \mathrm{Lie}^\circ_{A_\epsilon/R,2}) \otimes_R \hat{I};$$

the Hom space is a locally free R-module of rank $(s_\epsilon - j_{\epsilon,2})r_\epsilon$.

At Step $\alpha = 1, \dots, \epsilon$, we observe that condition (i) of the moduli problem $\underline{Y}_j^\circ$ implies $\phi_{\epsilon-\alpha+1,*,2}(\omega^\circ_{A^\vee_{\epsilon-\alpha+1}/R,2}) + \mathrm{Ker}(\phi^{\mathrm{dR}}_{\epsilon-\alpha,*,2})$ is an R-subbundle of $H_1^{\mathrm{dR}}(A_{\epsilon-\alpha}/R)^\circ_2$ of rank

$$s_{\epsilon-\alpha+1} - j_{\epsilon-\alpha+1,2} + j_{\epsilon-\alpha,2} = s_{\epsilon-\alpha} + (j_{\epsilon-\alpha+1,1} - j_{\epsilon-\alpha,2}) \quad \text{if } \alpha = 1, \dots, \epsilon-1, \tag{7.7.1}$$

and of rank $s_1 - j_{1,2}$ if $\alpha = \epsilon$. So $\phi^{\mathrm{cris}}_{\epsilon-\alpha+1,*,2}(\hat{\omega}_{\epsilon-\alpha+1,2}) + \mathrm{Ker}(\phi^{\mathrm{cris}}_{\epsilon-\alpha,*,2})$ is an $\hat{R}$-subbundle of $H_1^{\mathrm{cris}}(A_{\epsilon-\alpha}/\hat{R})^\circ_2$ of the same rank. The choices of the lifts $\hat{\omega}_{\epsilon-\alpha,2}$ form a torsor for the group

$$\mathrm{Hom}_R\big(\omega^\circ_{A^\vee_{\epsilon-\alpha}/R,2}/(\phi_{\epsilon-\alpha+1,*,2}(\omega^\circ_{A^\vee_{\epsilon-\alpha+1}/R,2}) + \mathrm{Ker}(\phi^{\mathrm{dR}}_{\epsilon-\alpha,*,2})), \mathrm{Lie}^\circ_{A_{\epsilon-\alpha}/R,2}\big) \otimes_R \hat{I}.$$

By (7.7.1), this Hom space is a locally free R-module of rank $(j_{\epsilon-\alpha+1,1} - j_{\epsilon-\alpha,2})r_{\epsilon-\alpha}$ if $\alpha = 1, \dots, \epsilon-1$ and of rank $(s_0 - (s_1 - j_{1,2}))r_0$ if $\alpha = \epsilon$. This implies that the contribution T_2 to the tangent space $T_{Y_j^\circ}$ at q_2 admits a filtration such that the subquotients are

$$\mathcal{H}om\big(\omega^\circ_{\mathcal{A}^\vee_{\epsilon-\alpha},2}/(\phi_{\epsilon-\alpha+1,*,2}(\omega^\circ_{\mathcal{A}^\vee_{\epsilon-\alpha+1},2}) + \mathrm{Ker}(\phi^{\mathrm{dR}}_{\epsilon-\alpha,*,2})), \mathrm{Lie}^\circ_{\mathcal{A}_{\epsilon-\alpha},2}\big)$$

where the $\mathcal{A}_{\epsilon-\alpha}$ are the universal abelian varieties and $\phi_{\epsilon+1,*,2}(\omega^\circ_{\mathcal{A}^\vee_{\epsilon+1},2})$ is interpreted as zero. In particular, T_2 is a locally free sheaf on Y_j° of rank

$$(s_\epsilon - j_{\epsilon,2})r_\epsilon + (s_0 - (s_1 - j_{1,2}))r_0 + \sum_{\alpha=1}^{\epsilon-1}(j_{\epsilon-\alpha+1,1} - j_{\epsilon-\alpha,2})r_{\epsilon-\alpha}$$
$$= (s_\epsilon - j_{\epsilon,2})r_\epsilon + j_{1,1}r_0 + \sum_{\alpha=1}^{\epsilon-1}(j_{\alpha+1,1} - j_{\alpha,2})r_\alpha. \quad (7.7.2)$$

We now look at the place q_1. By condition (ii), $\phi^{\mathrm{cris}}_{\alpha,*,1}$ when restricted to $\mathrm{Ker}(\psi^{\mathrm{cris}}_{\alpha+1,*,1})$ is a saturated injection of $\hat{R}$-bundles; and $\psi^{\mathrm{cris}}_{\alpha,*,1}$ when restricted to $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha-1,*,1})$ is also a saturated injection of $\hat{R}$-bundles. We first recall from the discussion in Notation 7.6 especially (7.6.1) that, when $\alpha = 1, \dots, \epsilon - 1$, $\omega^\circ_{A^\vee_\alpha/R,1}$ is the direct sum of

$$\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_\alpha/R,1} := \omega^\circ_{A^\vee_\alpha/R,1} \cap \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1}) \quad \text{and} \quad \omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_\alpha/R,1} := \omega^\circ_{A^\vee_\alpha/R,1} \cap \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha+1,*,1}),$$

which are locally free R-modules of rank r_ϵ and $r_\alpha - r_\epsilon$, respectively. Similarly, put

$$\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_\epsilon/R,1} := \omega^\circ_{A^\vee_\epsilon/R,1}, \quad \omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_\epsilon/R,1} := 0, \quad \text{and} \quad \omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_0/R,1} = \omega^\circ_{A^\vee_0/R,1} \cap \mathrm{Ker}(\psi^{\mathrm{dR}}_{1,*,1});$$

they have ranks r_ϵ, 0, and $r_0 - r_\epsilon$, respectively. We shall avoid talking about $\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_0/R,1}$ (as it does not make sense) but only psychologically understand it as the process that enlarges $\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_0/R,1}$ to $\omega^\circ_{A^\vee_0/R,1}$.

For $\alpha = 1, \dots, \epsilon$, the lift $\hat{\omega}_{\alpha,1}$ takes the form of $\hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha,1} \oplus \hat{\omega}^{\mathrm{Ker}\,\psi}_{\alpha,1}$, where the two direct summands are $\hat{R}$-subbundles of $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha,*,1})$ and of $\mathrm{Ker}(\psi^{\mathrm{cris}}_{\alpha+1,*,1})$, lifting $\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_\alpha/R,1}$ and $\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_\alpha/R,1}$, respectively. Whereas, the lift $\hat{\omega}_{0,1}$ contains the lift $\hat{\omega}^{\mathrm{Ker}\,\psi}_{0,1}$ of $\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_0/R,1}$ as an $\hat{R}$-subbundle of $\mathrm{Ker}(\psi^{\mathrm{cris}}_{1,*,1})$. Now the compatibility conditions $\phi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}_{\alpha,1}) \subseteq \hat{\omega}_{\alpha-1,1}$ and $\psi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}_{\alpha-1,1}) \subseteq \hat{\omega}_{\alpha,1}$ together with the condition (c) are equivalent to

$$\phi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}^{\mathrm{Ker}\,\psi}_{\alpha,1}) \subseteq \hat{\omega}^{\mathrm{Ker}\,\psi}_{\alpha-1,1} \quad \text{and} \quad \psi^{\mathrm{cris}}_{\alpha,*,1}(\hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha-1,1}) \subseteq \hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha,1}.$$

(The condition (c) on ranks of the quotients are also automatic.) In particular, the tangent space T_1 has three contributions, coming from the lifts $\hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha,1}$ (for $\alpha = 1, \dots, \epsilon$), from the lifts $\hat{\omega}^{\mathrm{Ker}\,\psi}_{\alpha,1}$ (for $\alpha = 0, \dots, \epsilon$), and from lifting $\omega^\circ_{A^\vee_0/R,1}$ to an $\hat{R}$-subbundle $\hat{\omega}_{0,1}$ of $H^{\mathrm{cris}}_1(A_0/\hat{R})^\circ_1$ containing $\hat{\omega}^{\mathrm{Ker}\,\psi}_{0,1}$. We shall use $T_1^{\mathrm{Ker}\,\phi}$, $T_1^{\mathrm{Ker}\,\psi}$, and $T_1^{\mathrm{Ker}\,\phi,0}$ to denote these three parts of the tangent space; and they will sit in an exact sequence

$$0 \to T_1^{\mathrm{Ker}\,\phi,0} \to T_1 \to T_1^{\mathrm{Ker}\,\phi} \oplus T_1^{\mathrm{Ker}\,\psi} \to 0. \quad (7.7.3)$$

We first determine the lifts $\hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha,1}$ for $\alpha = 1, \dots, \epsilon$. For $\hat{\omega}^{\mathrm{Ker}\,\phi}_{1,1}$, it lifts $\omega^{\circ,\mathrm{Ker}\,\phi}_{A^\vee_1/R,1}$ as an $\hat{R}$-subbundle of $H^{\mathrm{cris}}_1(A_1/\hat{R})^\circ_1$ of rank r_ϵ (with no further constraint). Then due to the rank constraint (and the injectivity of $\psi^{\mathrm{cris}}_{\alpha,*,1}$ when restricted to $\mathrm{Ker}(\phi^{\mathrm{cris}}_{\alpha+1,*,1})$), the lift $\hat{\omega}^{\mathrm{Ker}\,\phi}_{\alpha,1}$ for each $\alpha = 2, \dots, \epsilon$ is then forced to be equal to the image

$$\psi^{\mathrm{cris}}_{\alpha,*,1} \circ \cdots \circ \psi^{\mathrm{cris}}_{1,*,1}(\hat{\omega}^{\mathrm{Ker}\,\phi}_{1,1}).$$

So it suffices to consider the choices of the lift $\hat{\omega}^{\mathrm{Ker}\,\phi}_{1,1}$, which form a torsor for the group

$$\mathrm{Hom}_R(\omega^{\circ,\mathrm{Ker}\,\phi}_{A_1^\vee/R,1}, \mathrm{Ker}(\phi^{\mathrm{dR}}_{1,*,1})/\omega^{\circ,\mathrm{Ker}\,\phi}_{A_1^\vee/R,1}) \otimes_R \hat{I}.$$

This Hom space is a locally free R-module of rank

$$r_\epsilon(j_{1,1} - r_\epsilon). \tag{7.7.4}$$

It follows that the tangent space $T_1^{\mathrm{Ker}\,\phi}$ is simply just

$$\mathcal{H}om(\omega^{\circ,\mathrm{Ker}\,\phi}_{\mathcal{A}_0^\vee,1}, \mathrm{Ker}(\phi^{\mathrm{dR}}_{1,*,1})/\omega^{\circ,\mathrm{Ker}\,\phi}_{\mathcal{A}_0^\vee,1}).$$

We now determine the lifts $\hat{\omega}^{\mathrm{Ker}\,\psi}_{\alpha,1}$ for $\alpha = 0, \dots, \epsilon$ following the steps below:

Step 0: We start with putting $\hat{\omega}^{\mathrm{Ker}\,\psi}_{\epsilon,1} = 0$ because $\omega^{\circ,\mathrm{Ker}\,\psi}_{A_\epsilon^\vee/R,1}$ is,

Step(s) α: lift $\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_{\epsilon-\alpha}/R,1}$ to a subbundle $\hat{\omega}^{\mathrm{Ker}\,\psi}_{\epsilon-\alpha,1}$ of $\mathrm{Ker}(\psi^{\mathrm{cris}}_{\epsilon-\alpha+1,*,1})$ so that it contains $\phi^{\mathrm{cris}}_{\epsilon-\alpha+1}(\hat{\omega}^{\mathrm{Ker}\,\psi}_{\epsilon-\alpha+1,1})$,

Step ϵ: finally lift $\omega^{\circ,\mathrm{Ker}\,\psi}_{A_0^\vee/R,1}$ to a subbundle $\hat{\omega}^{\mathrm{Ker}\,\psi}_{0,1}$ of $\mathrm{Ker}(\psi^{\mathrm{cris}}_{1,*,1})$ so that it contains $\phi^{\mathrm{cris}}_{1,*,1}(\hat{\omega}^{\mathrm{Ker}\,\psi}_{1,1})$.

At Step $\alpha = 1, \dots, \epsilon$, the choices of the lifts $\hat{\omega}^{\mathrm{Ker}\,\psi}_{\epsilon-\alpha,1}$ form a torsor for the group

$$\mathrm{Hom}_R\big(\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_{\epsilon-\alpha}/R,1}\big/\phi_{\epsilon-\alpha+1,*,1}(\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_{\epsilon-\alpha+1}/R,1}), \mathrm{Ker}(\psi^{\mathrm{dR}}_{\epsilon-\alpha+1,*,1})\big/\omega^{\circ,\mathrm{Ker}\,\psi}_{A^\vee_{\epsilon-\alpha}/R,1}\big) \otimes_R \hat{I}.$$

This Hom space is a locally free R-module of rank

$$\big((r_{\epsilon-\alpha} - r_\epsilon) - (r_{\epsilon-\alpha+1} - r_\epsilon)\big)\big((n - j_{\epsilon-\alpha+1,1}) - (r_{\epsilon-\alpha} - r_\epsilon)\big).$$

This implies that the tangent space $T_1^{\mathrm{Ker}\,\psi}$ admits a filtration such that the subquotients are

$$\mathcal{H}om\big(\omega^{\circ,\mathrm{Ker}\,\psi}_{\mathcal{A}^\vee_{\epsilon-\alpha},1}\big/\phi_{\epsilon+1-\alpha,*,1}(\omega^{\circ,\mathrm{Ker}\,\psi}_{\mathcal{A}^\vee_{\epsilon+1-\alpha},1}), \mathrm{Ker}(\psi^{\mathrm{dR}}_{\epsilon+1-\alpha,*,1})/\omega^{\circ,\mathrm{Ker}\,\psi}_{\mathcal{A}^\vee_{\epsilon-\alpha},1}\big).$$

In particular, $T_1^{\mathrm{Ker}\,\psi}$ is a locally free sheaf on Y_j° of rank

$$\begin{aligned}\sum_{\alpha=1}^{\epsilon}&\big((r_{\epsilon-\alpha} - r_\epsilon) - (r_{\epsilon-\alpha+1} - r_\epsilon)\big)\big((n - j_{\epsilon-\alpha+1,1}) - (r_{\epsilon-\alpha} - r_\epsilon)\big)\\ &= \sum_{\alpha=0}^{\epsilon-1}(r_\alpha - r_{\alpha+1})(s_\alpha - j_{\alpha+1,1} + r_\epsilon).\end{aligned} \tag{7.7.5}$$

Finally, we discuss the $\hat{R}$-module $\hat{\omega}_{0,1}$ that lifts $\omega^{\circ}_{A_0^\vee/R,1}$ and contains $\hat{\omega}_{0,1}^{\mathrm{Ker}\,\psi}$ we obtained earlier. The lift is subject to one condition: $\hat{\omega}_{0,1} \subseteq (\psi^{\mathrm{cris}}_{1,*,1})^{-1}(\hat{\omega}_{1,1}^{\mathrm{Ker}\,\phi})$. So the choices of the lift form a torsor for the group

$$\mathrm{Hom}_R\big(\omega^{\circ}_{A_0^\vee/R,1}/\omega^{\circ,\mathrm{Ker}\,\psi}_{A_0^\vee/R,1},\, (\psi^{\mathrm{dR}}_{1,*,1})^{-1}(\omega^{\circ,\mathrm{Ker}\,\phi}_{A_1^\vee/R,1})/\omega^{\circ}_{A_0^\vee/R,1}\big) \otimes_R \hat{I}.$$

This implies that

$$T_1^{\mathrm{Ker}\,\phi,0} = \mathcal{H}om\big(\omega^{\circ}_{\mathcal{A}_0^\vee,1}/\omega^{\circ,\mathrm{Ker}\,\psi}_{\mathcal{A}_0^\vee,1},\, (\psi^{\mathrm{dR}}_{1,*,1})^{-1}(\omega^{\circ,\mathrm{Ker}\,\phi}_{\mathcal{A}_1^\vee,1})/\omega^{\circ}_{\mathcal{A}_0^\vee,1}\big),$$

which is locally free of rank

$$\big(r_0-(r_0-r_\epsilon)\big)\big((r_\epsilon+n-j_{1,1})-r_0\big) = r_\epsilon(s_0+r_\epsilon-j_{1,1}). \tag{7.7.6}$$

To sum up, the tangent space $T_{Y_j^\circ}$, as the direct sum $T_1 \oplus T_2$ with T_1 sitting in the exact sequence (7.7.3), is a locally free sheaf of rank given by $(7.7.6)+(7.7.4)+(7.7.5)+(7.7.2)$, that is,

$$\begin{aligned}
&r_\epsilon(s_0+r_\epsilon-j_{1,1}) + r_\epsilon(j_{1,1}-r_\epsilon) + \sum_{\alpha=0}^{\epsilon-1}(r_\alpha-r_{\alpha+1})(s_\alpha-j_{\alpha+1,1}+r_\epsilon)\\
&\qquad\qquad\qquad + (s_\epsilon-j_{\epsilon,2})r_\epsilon + j_{1,1}r_0 + \sum_{\alpha=1}^{\epsilon-1}(j_{\alpha+1,1}-j_{\alpha,2})r_\alpha\\
&= r_\epsilon s_0 + \sum_{\alpha=0}^{\epsilon-1} r_\alpha(s_\alpha-j_{\alpha+1,1}+r_\epsilon) - \sum_{\alpha=1}^{\epsilon} r_\alpha(s_{\alpha-1}-j_{\alpha,1}+r_\epsilon)\\
&\qquad\qquad\qquad + (s_\epsilon-j_{\epsilon,2})r_\epsilon + j_{1,1}r_0 + \sum_{\alpha=1}^{\epsilon-1}(j_{\alpha+1,1}-j_{\alpha,2})r_\alpha\\
&= r_\epsilon s_0 + r_0(s_0-j_{1,1}+r_\epsilon) + r_\epsilon(s_{\epsilon-1}-j_{\epsilon,1}+r_\epsilon) + (s_\epsilon-j_{\epsilon,2})r_\epsilon + j_{1,1}r_0\\
&\qquad + \sum_{\alpha=1}^{\epsilon-1} r_\alpha\big((s_\alpha-j_{\alpha+1,1}+r_\epsilon)-(s_{\alpha-1}-j_{\alpha,1}+r_\epsilon)+(j_{\alpha+1,1}-j_{\alpha,2})\big).
\end{aligned}$$

One easily checks that the first line adds up to $r_\epsilon s_\epsilon + r_0 s_0$, and the second line cancels to zero. This concludes the proof. □

In the special case of $\delta = r$, each abelian variety A_α appearing in the moduli problem of Y_j is isogenous to A_ϵ, which is a certain abelian variety parameterized by the discrete Shimura variety $\mathrm{Sh}_{0,n}$ and is hence supersingular (by Remark 3.7). So in particular, the image $\mathrm{pr}_j(Y_j)$ in this case is contained in the supersingular locus of $\mathrm{Sh}_{r,s}$. In fact, the converse is also true.

Theorem 7.8. *Assume $\delta = r$. The supersingular locus of $\mathrm{Sh}_{r,s}$ is the union of all $\mathrm{pr}_j(Y_j)$.*

Proof. We say a finite torsion $W(\bar{\mathbb{F}}_p)$-module has *divisible sequence* $(a_1, a_2, \dots, a_\epsilon)$ with nonnegative integers $a_1 \le \cdots \le a_\epsilon$ if it is isomorphic to

$$(W(\bar{\mathbb{F}}_p)/p^\epsilon)^{\oplus a_1} \oplus (W(\bar{\mathbb{F}}_p)/p^{\epsilon-1})^{\oplus(a_2-a_1)} \oplus \cdots \oplus (W(\bar{\mathbb{F}}_p)/p)^{\oplus(a_\epsilon - a_{\epsilon-1})}.$$

The following is an elementary linear algebra fact, whose proof we omit.

Claim: If $M_1 \subseteq M_2$ are two torsion $W(\bar{\mathbb{F}}_p)$-modules with divisible sequences $(a_{1,i},\dots,a_{\epsilon,i})$ for $i = 1, 2$ respectively, then $a_{\alpha,1} \le a_{\alpha,2}$ for all $\alpha = 1,\dots,\epsilon$.

The proof of the theorem is similar to the proof of Proposition 4.14(3), which is a special case of this theorem. It suffices to look at the closed points of $\mathrm{Sh}_{r,s}$. Let $z = (\mathcal{A}_z, \lambda, \eta) \in \mathrm{Sh}_{r,s}(\bar{\mathbb{F}}_p)$ be a supersingular point. Consider

$$\mathbb{L}_{\mathbb{Q}} = (\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1[1/p])^{F^2=p} = \{a \in \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1[1/p] \mid F^2(a) = pa\}.$$

Since x is supersingular, $\mathbb{L}_{\mathbb{Q}}$ is a $\mathbb{Q}_{p^2}$-vector space of dimension n, and $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1[1/p]$ may be identified with the extension of scalars of $\mathbb{L}_{\mathbb{Q}}$ from $\mathbb{Q}_{p^2}$ to $W(\bar{\mathbb{F}}_p)[1/p]$. Put

$$\tilde{\mathcal{E}}^\circ_1 = (\mathbb{L}_{\mathbb{Q}} \cap \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1) \otimes_{\mathbb{Z}_{p^2}} W(\bar{\mathbb{F}}_p) \quad \text{and} \quad \tilde{\mathcal{E}}^\circ_2 = F(\tilde{\mathcal{E}}^\circ_1) = V(\tilde{\mathcal{E}}^\circ_1) \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2.$$

Then we have

$$\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i/\tilde{\mathcal{E}}_i \simeq (W(\bar{\mathbb{F}}_p)/p^\epsilon)^{\oplus j_{1,i}} \oplus (W(\bar{\mathbb{F}}_p)/p^{\epsilon-1})^{\oplus(j_{2,i}-j_{1,i})} \oplus \cdots \cdots \oplus (W(\bar{\mathbb{F}}_p)/p)^{\oplus(j_{\epsilon,i}-j_{\epsilon-1,i})}, \quad (7.8.1)$$

for nondecreasing sequences $0 \le j_{1,i} \le j_{2,i} \le \cdots \le j_{\epsilon,i} \le n$ with $i = 1, 2$; in other words, $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i/\tilde{\mathcal{E}}_i$ has divisible sequence $(j_{1,i}, \dots, j_{\epsilon,i})$. Without loss of generality, we assume that $j_{1,1}$ and $j_{1,2}$ are not both zero. The essential part of the proof consists of checking the sequence of inequalities

$$0 \le j_{1,1} < j_{1,2} < j_{2,1} < j_{2,2} < \cdots < j_{\epsilon,1} < j_{\epsilon,2} \le n. \quad (7.8.2)$$

We first prove (7.8.2) with all strict inequalities replaced by nonstrict ones. Indeed, the obvious inclusion $F(\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i) \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_{3-i}$ implies that

$$F(\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1/\tilde{\mathcal{E}}_1) = F(\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1)/\tilde{\mathcal{E}}_2 \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2/\tilde{\mathcal{E}}_2, \quad \text{and}$$
$$F(\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2/\tilde{\mathcal{E}}_2) = F(\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2)/p\tilde{\mathcal{E}}_1 \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1/p\tilde{\mathcal{E}}_1.$$

By (7.8.1), the first inclusion embeds a torsion $W(\bar{\mathbb{F}}_p)$-module with divisible sequence $(j_{1,1}, \dots, j_{\epsilon,1})$ into a torsion $W(\bar{\mathbb{F}}_p)$-module with divisible sequence $(j_{1,2}, \dots, j_{\epsilon,2})$. The Claim above implies that $j_{\alpha,1} \le j_{\alpha,2}$ for all $\alpha = 1, \dots, \epsilon$. Similarly, by (7.8.1), the second inclusion embeds a torsion $W(\bar{\mathbb{F}}_p)$-module with divisible sequence $(j_{1,2}, \dots, j_{\epsilon,2})$ into a torsion $W(\bar{\mathbb{F}}_p)$-module with divisible sequence $(j_{1,1}, \dots, j_{\epsilon,1}, n)$. The Claim above implies that $j_{\alpha,2} \le j_{\alpha+1,1}$ for all $\alpha = 1, \dots, \epsilon - 1$, and $j_{\epsilon,2} \le n$.

We now use the construction of $\mathbb{L}_{\mathbb{Q}}$ to show the strict inequalities in (7.8.2). Suppose first that $j_{\alpha,1} = j_{\alpha,2}$ for some $\alpha = 1, \dots, \epsilon$. Then it follows that the maps

$$F,\ V : \Big(p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \frac{1}{p}\tilde{\mathcal{E}}^\circ_1\Big) + \tilde{\mathcal{E}}^\circ_1 \to \Big(p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2 \cap \frac{1}{p}\tilde{\mathcal{E}}^\circ_2\Big) + \tilde{\mathcal{E}}^\circ_2 \qquad (7.8.3)$$

are both isomorphisms (due to an easy length computation as $\tilde{\mathcal{E}}^\circ_2 = F(\tilde{\mathcal{E}}^\circ_1) = V(\tilde{\mathcal{E}}^\circ_1)$). By the definition of $\mathbb{L}_{\mathbb{Q}}$ and $\tilde{\mathcal{E}}^\circ_1$, we must have

$$\Big(\Big(p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \frac{1}{p}\tilde{\mathcal{E}}^\circ_1\Big) + \tilde{\mathcal{E}}^\circ_1\Big)^{F=V} \subseteq \mathbb{L}_{\mathbb{Q}} \cap \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \subseteq \tilde{\mathcal{E}}^\circ_1.$$

But this is absurd because the isomorphisms (7.8.3) implies by Hilbert's Theorem 90 that the left hand side above generates the source of (7.8.3), which is clearly not contained in $\tilde{\mathcal{E}}^\circ_1$.

Similarly, suppose that $j_{\alpha,2} = j_{\alpha+1,1}$ for some $\alpha = 1, \dots, \epsilon - 1$. Then the following morphisms are isomorphisms

$$F, V : \Big(p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2 \cap \frac{1}{p}\tilde{\mathcal{E}}^\circ_2\Big) + \tilde{\mathcal{E}}^\circ_2 \to (p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \tilde{\mathcal{E}}^\circ_1) + p\tilde{\mathcal{E}}^\circ_1, \qquad (7.8.4)$$

since $p\tilde{\mathcal{E}}^\circ_1 = F(\tilde{\mathcal{E}}^\circ_2) = V(\tilde{\mathcal{E}}^\circ_2)$ and for length reasons. By the definition of $\mathbb{L}_{\mathbb{Q}}$ and $\tilde{\mathcal{E}}^\circ_1$,

$$\big((p^{\epsilon-\alpha}\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \tilde{\mathcal{E}}^\circ_1) + p\tilde{\mathcal{E}}^\circ_1\big)^{F^{-1}=V^{-1}} \subseteq \mathbb{L}_{\mathbb{Q}} \cap p\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \subseteq p\tilde{\mathcal{E}}^\circ_1.$$

(Note that $\epsilon - \alpha \geq 1$ now.) But this is absurd because the isomorphisms (7.8.4) imply by Hilbert's Theorem 90 that the left hand side above generates the target of (7.8.4), which is clearly not contained in $p\tilde{\mathcal{E}}^\circ_1$.

Summing up, we have proved the strict inequalities (7.8.2). So the $j_{\alpha,i}$ define a $\boldsymbol{j}$ as in the beginning of Section 7.1. We now construct a point of $Y_{\boldsymbol{j}}$ which maps to the point $z \in \mathrm{Sh}_{r,s}$. Put

$$\tilde{\mathcal{E}}_{\alpha,1} := \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_1 \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_1 \quad \text{and} \quad \tilde{\mathcal{E}}_{\alpha,2} := \tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_2 \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_2. \qquad (7.8.5)$$

Using the exact construction in Section 7.2, we get the sequence of isogenies of abelian varieties

$$A_\epsilon \underset{\psi_\epsilon}{\overset{\phi_\epsilon}{\rightleftarrows}} A_{\epsilon-1} \underset{\psi_{\epsilon-1}}{\overset{\phi_{\epsilon-1}}{\rightleftarrows}} \cdots \underset{\psi_1}{\overset{\phi_1}{\rightleftarrows}} A_0 = \mathcal{A}_z,$$

such that A_α together with the induced polarization λ_α and the tame level structure η_α gives an $\bar{\mathbb{F}}_p$-point of $\mathrm{Sh}_{r_\alpha,s_\alpha}$, and $\tilde{\mathcal{D}}(A_\alpha)^\circ_i = \tilde{\mathcal{E}}_{\alpha,i}$ for all α and $i = 1, 2$.

Conditions (2)–(5) of Definition 7.4 easily follow from the description of the quotients $\tilde{\mathcal{D}}(\mathcal{A}_z)^\circ_i/\tilde{\mathcal{E}}_i$ in (7.8.1). Condition (6) of Definition 7.4 is equivalent to

$$p\tilde{\mathcal{D}}(A_{\alpha-1})^\circ_2 \subseteq V(\tilde{\mathcal{D}}(A_\alpha)^\circ_1).$$

By the construction of these Dieudonné modules in (7.8.5), this is equivalent to

$$p\Big(\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \cap \frac{1}{p^{\epsilon-\alpha+1}}\tilde{\mathcal{E}}_2\Big) \subseteq V\Big(\tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_1\Big).$$

But this follows from $p\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \subseteq V\tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ$ and $\tilde{\mathcal{E}}_2 = V\tilde{\mathcal{E}}_1$. Condition (7) of Definition 7.4 is equivalent to $\omega^\circ_{A^\vee_\alpha/\bar{\mathbb{F}}_p,1} \cap \mathrm{Ker}(\phi^{\mathrm{dR}}_{\alpha,*,1})$ having dimension r_ϵ, which is zero in our case. Translating it into the language of Dieudonné modules, this is equivalent to

$$V\tilde{\mathcal{D}}(A_\alpha)_2^\circ \cap p\tilde{\mathcal{D}}(A_{\alpha-1})_1^\circ = p\tilde{\mathcal{D}}(A_\alpha)_1^\circ.$$

By the construction of these Dieudonné module in (7.8.5), this is equivalent to

$$\Big(V\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \cap \frac{1}{p^{\epsilon-\alpha}}V\tilde{\mathcal{E}}_2\Big) \cap \Big(p\tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_1\Big) = p\tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ \cap \frac{1}{p^{\epsilon-\alpha-1}}\tilde{\mathcal{E}}_1,$$

which follows from observing that $V\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \supseteq p\tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ$ and $V\tilde{\mathcal{E}}_2 = p\tilde{\mathcal{E}}_1$. Condition (8) of Definition 7.4 is equivalent to $\omega^\circ_{A^\vee_{\alpha-1}/\bar{\mathbb{F}}_p,1} \subseteq \mathrm{Ker}(\psi^{\mathrm{dR}}_{\alpha,*,1})$ (note that $r_\epsilon = 0$ in our case). Translating it into the language of Dieudonné modules and using (7.8.5), this is equivalent to

$$V\tilde{\mathcal{D}}(A_{\alpha-1})_2^\circ \subseteq \tilde{\mathcal{D}}(A_\alpha)_1^\circ, \text{ or equivalently,}$$
$$V\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \cap \frac{1}{p^{\epsilon-\alpha+1}}V\tilde{\mathcal{E}}_2 \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ \cap \frac{1}{p^{\epsilon-\alpha}}\tilde{\mathcal{E}}_1,$$

which follows from observing that $V\tilde{\mathcal{D}}(\mathcal{A}_z)_2^\circ \subseteq \tilde{\mathcal{D}}(\mathcal{A}_z)_1^\circ$ and $V\tilde{\mathcal{E}}_2 = p\tilde{\mathcal{E}}_1$. This concludes the proof. □

Conjecture 7.9. *The varieties Y_j together with the natural morphisms to* $\mathrm{Sh}_{r-\delta,s+\delta}$ *and* $\mathrm{Sh}_{r,s}$ *satisfy condition* (3) *of Conjecture 2.12. Moreover, the union of the images of Y_j in* $\mathrm{Sh}_{r,s}$ *is the closure of the locus where the Newton polygon of the universal abelian variety has slopes* 0 *and* 1 *each with multiplicity* $2(r-\delta)n$, *and slope* $\frac{1}{2}$ *with multiplicity* $2(n-2r+2\delta)n$.

This conjecture in the case of $r = \delta = 1$ was proved in Theorem 4.18.

Appendix A: An explicit formula in the local spherical Hecke algebra for GL_n

In this appendix, let F be a local field with ring of integers $\mathcal{O}$, $\varpi \in \mathcal{O}$ be a uniformizer, $\mathbb{F} = \mathcal{O}/\varpi\mathcal{O}$ and $q = \#\mathbb{F}$. Fix an integer $n \geq 1$. We consider the spherical Hecke algebra $\mathscr{H}_K = \mathbb{Z}[K \backslash \mathrm{GL}_n(F)/K]$ with $K = \mathrm{GL}_n(\mathcal{O})$. Here, the product of two double cosets $u = KxK$ and $v = KyK$ in $\mathcal{H}_K$ is defined as

$$u \cdot v = \sum_w m(u, v; w)w,^{19} \tag{A.0.1}$$

where the sum runs through all the double cosets $w = KzK$ contained in $KxKyK$, and the coefficient $m(u, v; w) \in \mathbb{Z}$ is determined as follows: If $KxK = \coprod_{i\in I} x_i K$ and $KyK = \coprod_{j\in J} y_j K$, then

$$m(u, v; w) = \#\{(i, j) \in I \times J \mid x_i y_j K = zK \text{ for a fixed element } z \text{ in } w\}. \tag{A.0.2}$$

By the theory of elementary divisors, all double cosets KxK are of the form

$$T(a_1, \dots, a_n) := K \operatorname{Diag}(\varpi^{a_1}, \dots, \varpi^{a_n})K \quad \text{for } a_i \in \mathbb{Z} \text{ with } a_1 \geq a_2 \geq \cdots \geq a_n.$$

They form a $\mathbb{Z}$-basis of $\mathcal{H}_K$. We put

$$T^{(r)} = T(\underbrace{1, \dots, 1}_{r}, \underbrace{0, \dots, 0}_{n-r}) \qquad \text{for } 0 \leq r \leq n,$$

$$R^{(r,s)} = T(\underbrace{2, \dots, 2}_{r}, \underbrace{1, \dots, 1}_{s-r}, \underbrace{0, \dots, 0}_{n-s}) \quad \text{for } 0 \leq r \leq s \leq n.$$

In particular, $R^{(0,s)} = T^{(s)}$ and $T^{(0)} = [K]$.

Because of the lack of references, we include a proof of the following:

Proposition A.1. *For* $1 \leq r \leq n$, *let*

$$\binom{n}{r}_q = \frac{(q^n - 1)(q^{n-1} - 1) \cdots (q^{n-r+1} - 1)}{(q - 1)(q^2 - 1) \cdots (q^r - 1)} \tag{A.1.1}$$

be the Gaussian binomial coefficients, and put $\binom{n}{0}_q = 1$. *Then for* $0 \leq r \leq s \leq n$,

$$T^{(r)} T^{(s)} = \sum_{i=0}^{\min\{r,n-s\}} \binom{s-r+2i}{i}_q R^{(r-i,s+i)}.$$

Proof. We fix a set of representatives $\tilde{\mathbb{F}} \subseteq \mathcal{O}$ of $\mathbb{F} = \mathcal{O}/\varpi\mathcal{O}$ which contains 0. Then we have $T^{(r)} = \coprod_{x\in\mathcal{S}(n,r)} xK$, where $\mathcal{S}(n, r)$ is the set of $n \times n$ matrices $x = (x_{i,j})_{1\leq i,j\leq n}$ such that

- r of the diagonal entries are ϖ and the remaining $n - r$ ones are 1;
- if $i \neq j$, then $x_{i,j} = 0$ unless $i > j$, $x_{i,i} = 1$ and $x_{j,j} = \varpi$, in which case $x_{i,j}$ can take any values in $\tilde{\mathbb{F}}$.

[19] We may also view elements of $\mathcal{H}_K$ as $\mathbb{Z}$-valued locally constant and compactly supported functions on $\mathrm{GL}_n(F)$ which are bi-invariant under K, and define the product of $f, g \in \mathcal{H}_K$ as $(f * g)(x) = \int_{\mathrm{GL}_n(F)} f(y)g(y^{-1}x)\,dy$, where dy means the unique bi-invariant Haar measure on $\mathrm{GL}_n(F)$ with $\int_K dy = 1$. For the equivalence between these two definitions, see [Gross 1998, p. 4].

For instance, the set $\mathcal{S}(3, 2)$ consists of matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ x_{2,1} & \varpi & 0 \\ x_{3,1} & 0 & \varpi \end{pmatrix}, \quad \begin{pmatrix} \varpi & 0 & 0 \\ 0 & 1 & 0 \\ 0 & x_{3,2} & \varpi \end{pmatrix}, \quad \begin{pmatrix} \varpi & 0 & 0 \\ 0 & \varpi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $x_{2,1}, x_{3,1}, x_{3,2} \in \tilde{\mathbb{F}}$. We have a similar decomposition $T^{(s)} = \coprod_{y \in \mathcal{S}(n,s)} yK$. We write $T^{(r)}T^{(s)}$ as a linear combination of $T(a_1, \dots, a_n)$ with $a_i \in \mathbb{Z}$ and $a_1 \geq \dots \geq a_n$. By looking at the diagonal entries of xy, we see easily that only $R^{(r-i,s+i)}$ with $0 \leq i \leq \min\{r, n-s\}$ have nonzero coefficients, namely, we have

$$T^{(r)}T^{(s)} = \sum_{i=0}^{\min\{r,n-s\}} C^{(r,s)}(n, i) R^{(r-i,s+i)} \qquad \text{for some } C^{(r,s)}(n, i) \in \mathbb{Z}.$$

By (A.0.1), $C^{(r,s)}(n, i)$ is the number of pairs $(x, y) \in \mathcal{S}(n, r) \times \mathcal{S}(n, s)$ such that

$$xyK = \operatorname{Diag}(\underbrace{\varpi^2, \dots, \varpi^2}_{r-i}, \underbrace{\varpi, \dots, \varpi}_{s-r+2i}, \underbrace{1, \dots, 1}_{n-s-i})K.$$

In this case, x and y must be of the form

$$x = \begin{pmatrix} \varpi I_{r-i} & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & I_{n-s-i} \end{pmatrix}, \quad y = \begin{pmatrix} \varpi I_{r-i} & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & I_{n-s-i} \end{pmatrix},$$

where I_k denotes the $k \times k$ identity matrix, and $A \in \mathcal{S}(s - r + 2i, i)$ and $B \in \mathcal{S}(s - r + 2i, s - r + i)$ satisfy $AB \cdot \mathrm{GL}_{s-r+2i}(\mathcal{O}) = \varpi I_{s-r+2i} \mathrm{GL}_{s-r+2i}(\mathcal{O})$. By (A.0.1), we see that $C^{(r,s)}(n, i) = C^{(i,s-r+i)}(s-r+2i, i)$. Therefore, one is reduced to proving the following lemma, which is a special case of our proposition. □

Lemma A.2. *Under the notation and hypothesis of Proposition A.1, assume moreover that $n = r + s$. Then the coefficient of $R^{(0,n)}$ in the product $T^{(r)}T^{(s)}$ is $\binom{n}{r}_q$.*

Proof. We induct on $n \geq 1$. The case $n = 1$ is trivial. We assume thus $n > 1$, and that the statement is true when n is replaced by $n - 1$. The case of $r = 0$ being trivial, we may assume that $r \geq 1$. We say a pair $(x, y) \in \mathcal{S}(n, r) \times \mathcal{S}(n, n - r)$ is admissible if $xyK = \varpi I_n K$. We have to show that the number of admissible pairs is equal to $\binom{n}{r}_q$. Let (x, y) be an admissible pair. Denote by I (resp. by J) the set integers $1 \leq i \leq n$ such that $x_{i,i} = \varpi$ (resp. $y_{i,i} = \varpi$). Note that (x, y) being admissible implies that $J = \{1, \dots, n\} \setminus I$.

Assume first that $x_{1,1} = 1$. Then x and y must be of the form $x = \left(\begin{smallmatrix} 1 & 0 \\ * & A \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} \varpi & 0 \\ 0 & B \end{smallmatrix}\right)$ where $(A, B) \in \mathcal{S}(n-1, r) \times \mathcal{S}(n-1, n-1-r)$ admissible. Note that $xyK = \varpi I_n K$ always hold. We have $x_{i,1} = 0$ for $i \notin I$, and $x_{i,1}$ can take any values in $\mathbb{F}$ for $i \in I$. Therefore, the number of admissible pairs (x, y) with $x_{1,1} = 1$ is equal to $q^{\#I} = q^r$ times that of the admissible (A, B). The latter is equal to $\binom{n-1}{r}_q$ by the induction hypothesis.

Consider now the case $x_{1,1} = \varpi$. One can write $x = \left(\begin{smallmatrix} \varpi & 0 \\ 0 & A \end{smallmatrix}\right)$, and $y = \left(\begin{smallmatrix} 1 & 0 \\ * & B \end{smallmatrix}\right)$ with $(A, B) \in \mathcal{S}(n-1, r-1) \times \mathcal{S}(n-1, n-r)$ admissible. Put $z = xy$. Then an easy computation shows that $z_{j,1} = y_{j,1}$ if $j \in J$, and $z_{j,1} = 0$ if $j \notin J$. Hence, $xyK = \varpi I_n K$ forces that $y_{j,1} = 0$ for all $j > 1$. Therefore, the number of admissible (x, y) in this case is equal to that of the admissible (A, B), which is $\binom{n-1}{r-1}_q$ by the induction hypothesis. The lemma now follows immediately from the equality

$$\binom{n}{r}_q = q^r \binom{n-1}{r}_q + \binom{n-1}{r-1}_q. \qquad \square$$

Appendix B: A determinant formula

In this appendix, we prove the following:

Theorem B.1. *Let $\alpha_1, \ldots, \alpha_n$ be n indeterminates. For $i = 1, \ldots, n$, let s_i denote the i-th elementary symmetric polynomial in the α, and $s_0 = 1$ by convention. Let q be another indeterminate. We put $q_r = q^{r-1} + q^{r-3} + \cdots + q^{1-r}$. Consider the matrix $M_n(q) = (m_{i,j})$ given as follows*:

$$m_{i,j} = \begin{cases} \sum_{\delta=0}^{\min\{i-1,n-j\}} q_{n+i-j-2\delta} s_{j-i+\delta} s_{n-\delta} & \text{if } i \le j; \\ \sum_{\delta=0}^{\min\{j-1,n-i\}} q_{n+j-i-2\delta} s_\delta s_{n+j-i+\delta} & \text{if } i > j. \end{cases}$$

Then we have

$$\det(M_n(q)) = \alpha_1 \cdots \alpha_n \prod_{i \ne j} \Big(q\alpha_i - \frac{1}{q} \alpha_j \Big).$$

Proof. Let $N_n(q)$ be the resultant matrix of the polynomials $f(x) = \prod_{i=1}^n (x + q^{-1}\alpha_i)$ and $g(x) = \prod_{i=1}^n (x + q\alpha_i)$, that is, $N_n(q)$ is the $2n \times 2n$ matrix given by

$$N_n(q) = \begin{pmatrix} s_0 & q^{-1}s_1 & q^{-2}s_2 & \cdots & q^{1-n}s_{n-1} & q^{-n}s_n & 0 & \cdots & 0 \\ 0 & s_0 & q^{-1}s_1 & \cdots & q^{2-n}s_{n-2} & q^{1-n}s_{n-1} & q^{-n}s_n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_0 & q^{-1}s_1 & q^{-2}s_2 & \cdots & q^{-n}s_n \\ s_0 & qs_1 & q^2s_2 & \cdots & q^{n-1}s_{n-1} & q^n s_n & 0 & \cdots & 0 \\ 0 & s_0 & qs_1 & \cdots & q^{n-2}s_{n-2} & q^{n-1}s_{n-1} & q^n s_n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_0 & qs_1 & q^2s_2 & \cdots & q^n s_n \end{pmatrix}.$$

It is well known that $\det(N_n(q)) = \prod_{i,j}(-q^{-1}\alpha_i + q\alpha_j)$. Thus it suffices to show that $\det(N_n(q)) = (q - q^{-1})^n \det(M_n(q))$.

We first make the following row operations on $N_n(q)$: subtract row i from row $n+i$ for all $i = 1, \ldots, n$. We obtain a matrix whose first column is all 0 except the first entry being 1; moreover, one can take out a factor $(q-q^{-1})$ from row $n+1,\dots,2n$. Let $N'_n(q)$ be the right lower $(2n-1)\times(2n-1)$ submatrix of the remaining matrix. Then we have

$$N'_n(q)=\begin{pmatrix} s_0 & q^{-1}s_1 & q^{-2}s_2 & \cdots & q^{1-n}s_{n-1} & q^{-n}s_n & 0 & \cdots & 0\\ 0 & s_0 & q^{-1}s_1 & \cdots & q^{2-n}s_{n-2} & q^{1-n}s_{n-1} & q^{-n}s_n & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & s_0 & q^{-1}s_1 & q^{-2}s_2 & \cdots & q^{-n}s_n\\ q_1s_1 & q_2s_2 & q_3s_3 & \cdots & q_{n-1}s_{n-1} & q_ns_n & 0 & \cdots & 0\\ 0 & q_1s_1 & q_2s_2 & \cdots & q_{n-2}s_{n-2} & q_{n-1}s_{n-1} & q_ns_n & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 0 & q_1s_1 & q_2s_2 & \cdots & q_ns_n \end{pmatrix}$$

with $\det(N_n(q))=(q-q^{-1})^n\det(N'_n(q))$. Thus we are reduced to proving that $\det(N'_n(q))=\det(M_n(q))$. Consider the $(2n-1)\times(2n-1)$ matrix $R=\left(\begin{smallmatrix} I_{n-1} & 0\\ C & D\end{smallmatrix}\right)$ with the lower $n\times(2n-1)$ submatrix given by

$$\begin{pmatrix} C & D\end{pmatrix}=\begin{pmatrix} -q_1s_1 & -q_2s_2 & \cdots & -q_{n-1}s_{n-1} & 1 & q^{-1}s_1 & q^{-2}s_2 & \cdots & q^{2-n}s_{n-2} & q^{1-n}s_{n-1}\\ 0 & -q_1s_1 & \cdots & -q_{n-2}s_{n-2} & 0 & 1 & q^{-1}s_1 & \cdots & q^{3-n}s_{n-3} & q^{2-n}s_{n-2}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & -q_1s_1 & 0 & 0 & 0 & \cdots & 1 & q^{-1}s_1\\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

By a careful computation, one verifies without difficulty that $RN'_n(q)=\left(\begin{smallmatrix} U & *\\ 0 & M_n(q)\end{smallmatrix}\right)$, where U is an $(n-1)\times(n-1)$-upper triangular matrix with all diagonal entries equal to 1. Note that $\det(R)=\det(D)=\det(U)=1$, and it follows immediately that $\det(N'_n(q))=\det(M_n(q))$. □

Acknowledgments. We thank Xinwen Zhu for his insights into Conjecture 2.12 and allowing us to include them in this current paper. We also thank Gaëtan Chenevier, Matthew Emerton, Robert Kottwitz, Sug-Woo Shin and Jacques Tilouine for extremely helpful discussions. We thank the anonymous referees for their careful reading of the paper and helpful comments. We thank the Fields Institute for hosting the thematic program, during which this collaboration started. Helm thanks University of Chicago and University of California at Irvine for hospitality when he visited. Tian thanks the hospitality of Institut de Hautes Études Scientifiques where part of this work was written. Tian thanks University of California at Irvine and University of Connecticut at Storrs for hospitality when he visited. Helm was partially supported by NSF grant DMS–1161582. Tian was supported in part by the National Natural Science Foundation of China (No. 11321101). Xiao was partially supported by Simons Collaboration Grant #278433 and NSF Grant DMS–1502147.

References

[Caraiani 2012] A. Caraiani, "Local-global compatibility and the action of monodromy on nearby cycles", *Duke Math. J.* **161**:12 (2012), 2311–2413. MR Zbl

[Di Francesco 1997] P. Di Francesco, "SU(N) meander determinants", *J. Math. Phys.* **38**:11 (1997), 5905–5943. MR Zbl

[Fontaine et al. 2013] B. Fontaine, J. Kamnitzer, and G. Kuperberg, "Buildings, spiders, and geometric Satake", *Compos. Math.* **149**:11 (2013), 1871–1912. MR Zbl

[Fulton 1998] W. Fulton, *Intersection theory*, Second ed., Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics **2**, Springer, 1998. MR Zbl

[Görtz and He 2015] U. Görtz and X. He, "Basic loci of Coxeter type in Shimura varieties", *Camb. J. Math.* **3**:3 (2015), 323–353. MR Zbl

[Gross 1998] B. H. Gross, "On the Satake isomorphism", pp. 223–237 in *Galois representations in arithmetic algebraic geometry* (Durham, 1996), edited by A. J. Scholl and R. L. Taylor, London Math. Soc. Lecture Note Ser. **254**, Cambridge Univ. Press, 1998. MR Zbl

[Grothendieck 1974] A. Grothendieck, *Groupes de Barsotti–Tate et cristaux de Dieudonné* (Séminaire de Mathématiques Supérieures, No. 45 Été, 1970), Les Presses de l'Université de Montréal, Montreal, Quebec, 1974. MR Zbl

[Harris and Taylor 2001] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies **151**, Princeton University Press, Princeton, NJ, 2001. With an appendix by Vladimir G. Berkovich. MR Zbl

[Helm 2010] D. Helm, "Towards a geometric Jacquet–Langlands correspondence for unitary Shimura varieties", *Duke Math. J.* **155**:3 (2010), 483–518. MR Zbl

[Helm 2012] D. Helm, "A geometric Jacquet–Langlands correspondence for $U(2)$ Shimura varieties", *Israel J. Math.* **187** (2012), 37–80. MR Zbl

[Howard and Pappas 2014] B. Howard and G. Pappas, "On the supersingular locus of the GU(2, 2) Shimura variety", *Algebra Number Theory* **8**:7 (2014), 1659–1699. MR Zbl

[Jacquet and Shalika 1981] H. Jacquet and J. A. Shalika, "On Euler products and the classification of automorphic representations. I", *Amer. J. Math.* **103**:3 (1981), 499–558. MR Zbl

[Kisin 2010] M. Kisin, "Integral models for Shimura varieties of abelian type", *J. Amer. Math. Soc.* **23**:4 (2010), 967–1012. MR Zbl

[Kottwitz 1992a] R. E. Kottwitz, "On the λ-adic representations associated to some simple Shimura varieties", *Invent. Math.* **108**:3 (1992), 653–665. MR Zbl

[Kottwitz 1992b] R. E. Kottwitz, "Points on some Shimura varieties over finite fields", *J. Amer. Math. Soc.* **5**:2 (1992), 373–444. MR Zbl

[Lan 2013] K.-W. Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, London Mathematical Society Monographs Series **36**, Princeton University Press, Princeton, NJ, 2013. MR Zbl

[Mazur and Messing 1974] B. Mazur and W. Messing, *Universal extensions and one dimensional crystalline cohomology*, Lecture Notes in Mathematics **370**, Springer, 1974. MR Zbl

[Milne 1990] J. S. Milne, "Canonical models of (mixed) Shimura varieties and automorphic vector bundles", pp. 283–414 in *Automorphic forms, Shimura varieties, and L-functions* (Ann Arbor, MI, 1988), vol. I, edited by L. Clozel and J. S. Milne, Perspect. Math. **10**, Academic Press, Boston, 1990. MR Zbl

[Mumford 2008] D. Mumford, *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics **5**, Hindustan Book Agency, New Delhi, 2008. MR Zbl

[Pappas and Rapoport 2005] G. Pappas and M. Rapoport, "Local models in the ramified case. II. Splitting models", *Duke Math. J.* **127**:2 (2005), 193–250. MR Zbl

[Rapoport et al. 2014] M. Rapoport, U. Terstiege, and S. Wilson, "The supersingular locus of the Shimura variety for $\mathrm{GU}(1, n-1)$ over a ramified prime", *Math. Z.* **276**:3-4 (2014), 1165–1188. MR Zbl

[Reduzzi and Xiao 2017] D. A. Reduzzi and L. Xiao, "Partial Hasse invariants on splitting models of Hilbert modular varieties", *Ann. Sci. Éc. Norm. Supér.* (4) **50**:3 (2017), 579–607. MR Zbl

[Serre 1996] J.-P. Serre, "Two letters on quaternions and modular forms (mod p)", *Israel J. Math.* **95** (1996), 281–299. With introduction, appendix and references by R. Livné. MR

[SGA 4_2 1972] M. Artin, A. Grothendieck, and J. Verdier, *Théorie des topos et cohomologie étale des schémas. Tome 2* (Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964), Lecture Notes in Mathematics **270**, Springer, 1972. MR Zbl

[Shin 2014] S.-W. Shin, "On the cohomological base change for unitary similitude groups", (2014). appendix to W. Goldring, "Galois representations associated to holomorphic limits of discrete series I: Unitary Groups", *Compositio Math.* **150**:2 (2014), 191–228.

[Tate 1966] J. Tate, "Endomorphisms of abelian varieties over finite fields", *Invent. Math.* **2** (1966), 134–144. MR Zbl

[Tian and Xiao 2014] Y. Tian and L. Xiao, "Tate cycles on some quaternionic Shimura varieties mod p", 2014. arXiv

[Tian and Xiao 2016] Y. Tian and L. Xiao, "On Goren–Oort stratification for quaternionic Shimura varieties", *Compos. Math.* **152**:10 (2016), 2134–2220. MR Zbl

[Vollaard and Wedhorn 2011] I. Vollaard and T. Wedhorn, "The supersingular locus of the Shimura variety of $\mathrm{GU}(1, n-1)$ II", *Invent. Math.* **184**:3 (2011), 591–627. MR Zbl

[White 2012] P.-J. White, "Tempered automorphic representation of the unitary group", preprint, 2012, Available at http://preprints.ihes.fr/2011/M/M-11-17.pdf.

[Xiao and Zhu 2017] L. Xiao and X. Zhu, "Cycles on Shimura varieties via geometric Satake", 2017. arXiv

[Zhu 2017] X. Zhu, "Affine Grassmannians and the geometric Satake in mixed characteristic", *Ann. of Math.* (2) **185**:2 (2017), 403–492. MR Zbl

Communicated by Brian Conrad
Received 2015-11-17 Revised 2017-08-24 Accepted 2017-09-28

d.helm@imperial.ac.uk *Department of Mathematics, Imperial College London, South Kensington Campus, London, United Kingdom*

tian@math.uni-bonn.de *Mathematics Institute, University of Bonn, Bonn, Germany*

liang.xiao@uconn.edu *Department of Mathematics, University of Connecticut, Storrs, CT, United States*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2289

Complex conjugation and Shimura varieties

Don Blasius and Lucio Guerberoff

In this paper we study the action of complex conjugation on Shimura varieties and the problem of descending Shimura varieties to the maximal totally real field of the reflex field. We prove the existence of such a descent for many Shimura varieties whose associated adjoint group has certain factors of type A or D. This includes a large family of Shimura varieties of abelian type. Our considerations and constructions are carried out purely at the level of Shimura data and group theory.

1. Introduction

The goal of this paper is to analyze some aspects of complex conjugation acting on Shimura varieties. This topic has been studied for a long time by several authors, notably Shimura, Deligne, Langlands, Milne, Shih, and more recently Taylor. In general, given a Shimura variety $\mathrm{Sh}(G, X)$ defined by a Shimura datum (G, X), and any automorphism α of $\mathbb{C}$, Langlands [1979] conjectured that the conjugate variety $\alpha\,\mathrm{Sh}(G, X) = \mathrm{Sh}(G, X) \times_{\mathbb{C},\alpha} \mathbb{C}$ can be realized as a Shimura variety $\mathrm{Sh}({}^{\alpha}G, {}^{\alpha}X)$ for a very explicit pair $({}^{\alpha}G, {}^{\alpha}X)$. This has been proved by Milne [1983] (see also [Borovoĭ 1983; 1987; Milne 1999]). The case of $\alpha = c$ (complex conjugation) has, among other properties, the particularity that the pair $({}^{c}G, {}^{c}X)$ is very concrete. Namely, it can be identified with $(G, \overline{X})$, where $\overline{X}$ is obtained by composing the elements of x with complex conjugation on the Deligne torus $\mathbb{S}$. This simple description is hard to find in the literature, and hence, we include a proof of how it is deduced from the general constructions.

Assuming a few standard extra conditions on the Shimura datum (G, X), the reflex field E can be seen to be either totally real or a CM field. The Shimura variety has a canonical model $\mathrm{Sh}(G, X)_E$ over E, and the Hecke operators are defined over E as well. In this paper we investigate descent of these varieties to the maximal totally real subfield E^+ of E. The existence of such descent can be seen as a nice generalization of the useful fact that the field obtained by adjoining to $\mathbb{Q}$ the j-invariant of an order in an imaginary quadratic field has a real embedding. From now on, assume that E is CM. We show in many cases that $\mathrm{Sh}(G, X)$ has

MSC2010: primary 11G18; secondary 11G35, 11E57, 20G30.
Keywords: Shimura varieties, conjugation, models.

a model over E^+. Although the Hecke operators are not defined over E^+, they can nevertheless be characterized. The general framework for constructing such models comes from the construction of descent data arising from automorphisms of G of order 2 taking X to $\overline{X}$. Using the classification of (adjoint) Shimura data in terms of special nodes on Dynkin diagrams, our aim is to construct an involution of G that induces the opposition involution on the based root datum (or the Dynkin diagram). The construction we make follows from the classification of semisimple groups. The groups G with which we work are, roughly speaking, those for which the simple factors of G^{ad} are of classical type A or D, and satisfy an extra condition on the hermitian or skew-hermitian space defining them (see Definitions 4.1.4 and 4.2.1). For example, a factor of type A is attached to a hermitian space over a central division algebra D over a CM field K endowed with an involution of the second kind J. We show that, if there exists an opposition involution on these groups, then D must be either K or a quaternion division algebra, and the involution J is easily described. We carry out the construction of involutions if we assume the aforementioned extra condition, which in this case amounts to the existence of a basis of the underlying vector space such that the matrix of the hermitian form is diagonal with entries in K. In the quaternion algebra case, we can write $D = D_0 \otimes_F K$, where F is the maximal totally real subfield of K, and D_0 is a quaternion division algebra over F. We assume furthermore in this case that, if $D_{0,v}$ is not split for an embedding $v : F \hookrightarrow \mathbb{R}$, then the corresponding factor of $G^{\mathrm{ad}}_{\mathbb{R}}$ is compact. If $D = K$, the conditions in Definition 4.1.4 are automatically satisfied. For factors of type D, there is a similar scenario, although we only restrict to groups of type $D^{\mathbb{H}}$ as in the Appendix of [Milne and Shih 1981]. This encompasses a large family of Shimura varieties of abelian type. Under these assumptions, the existence of the involution on the group G follows from a concrete construction of involutions on each of the simple factors of G^{ad}, which are explicitly given in terms of simple algebras.

To give a flavor of the type of involutions constructed in the paper, suppose that $\mathrm{SU}(V, h)$ is a simple factor of type A, corresponding to a hermitian space (V, h) of dimension n over a CM field K. Let F be the maximal totally real subfield of K, and let ι be the nontrivial automorphism of K/F. We take an orthogonal basis $\{v_1, \ldots, v_n\}$ of V, and we let $I : V \to V$ be the ι-semilinear map obtained by applying ι to the coordinates of elements of V with respect to the given basis. Then the map $\theta : \mathrm{SU}(V, h) \to \mathrm{SU}(V, h)$ given by

$$\theta(g) = IgI$$

is an opposition involution.

We stress here that our methods are group-theoretic and we work purely at the level of Shimura data, in the sense that we do not directly make use of a moduli interpretation. However, the methods rely on Langlands conjugation of Shimura varieties, which in turn is proved using the moduli interpretation in terms of abelian varieties [Milne 1990, §II.9]. An interesting question would be to consider factors of type E_6, which is the only other type apart from A or $D^{\mathbb{H}}$ that contributes to the reflex field being CM instead of totally real. We plan to investigate this question in the future.

Let us describe the organization of the paper and outline the main argument. In Section 2, we start by recalling the general formalism of conjugation of Shimura varieties by an arbitrary automorphism of $\mathbb{C}$, we study the special case of complex conjugation explicitly, and we prove in this case that the conjugate Shimura datum is $(G, \overline{X})$, where $\overline{X}$ is the complex conjugate conjugacy class of X. We show (Theorem 2.3.1) that, if (G, X) is a Shimura datum and $\theta : G \to G$ is an involution such that $\theta(X) = \overline{X}$, then θ induces an isomorphism of algebraic varieties from the complex conjugate $c\,\mathrm{Sh}(G, X)$ to $\mathrm{Sh}(G, X)$, defined over the reflex field E, that constitutes a descent datum from E to E^+.

In Section 3, we recall some basic facts about root data and opposition involutions, and in Proposition 3.4.8, we lay the ground for the prototype of involutions $\theta : G \to G$ that we will construct. Roughly speaking, suppose that $T \subset G$ is a maximal torus of G, and $x \in X$ factors through $T_{\mathbb{R}}$. If $\theta : G \to G$ is an involution that preserves $T_{\mathbb{R}}$ and induces complex conjugation on the group of characters $X^*(T)$, then $\theta(x) = \bar{x}$ and thus $\theta(X) = \overline{X}$. This is basically the type of involution that we will construct, with some slight changes. Since we will make use of the explicit classification of semisimple groups, we need to work with either G^{der} or G^{ad}. We let G_i be the almost simple factors of G^{der}, and $\widetilde{G}_i$ be their simply connected covers, so that $\widetilde{G}_i = \mathrm{Res}_{F_i/\mathbb{Q}} H_i$, for certain groups H_i which are absolutely almost simple, simply connected, over a totally real field F_i. We recall the classification of these groups in Section 4, where we also construct opposition involutions on them preserving specific maximal tori S_i and inducing complex conjugation on their characters (for noncompact places v of F_i). We only do this for groups of type A or $D^{\mathbb{H}}$. These, together with type E_6, are the only ones that give a CM reflex field, as opposed to totally real. Furthermore, as noted above, we impose some extra conditions in order to construct the involutions. From the tori S_i, we get maximal tori $T' \subset G^{\mathrm{der}}$ and $T \subset G$, and an opposition involution $\theta' : G^{\mathrm{der}} \to G^{\mathrm{der}}$ preserving T'. As shown in Proposition 3.4.8, θ' extends uniquely to an involution on G. To show that $\theta(X) = \overline{X}$, we need to relate in some way the choice of our tori S_i, which is a priori unrelated to the Shimura datum, to the conjugacy class X. In Section 5, we show that there always exists $x \in X$ such that x^{ad} factors through the image of $T_{\mathbb{R}}$ in $G^{\mathrm{ad}}_{\mathbb{R}}$. This is all we need for Proposition 3.4.8. In Theorem 5.2.2, we state the existence of descent datum for Shimura varieties defined by groups (G, X) such that the simple factors of G^{ad} are of the type described in Section 4. We call these *strongly of type* $(AD^{\mathbb{H}})$. Finally, we also note that involutions inducing the desired descent datum on $\mathrm{Sh}(G, X)$ can be constructed whenever G is adjoint and there exists an opposition involution $\theta : G \to G$. This is always the case if G is quasisplit, for example.

The existence of the involutions constructed in this paper should have interesting applications, which will be explored in the future, for example, in the setting of integral models and the zeta function problem, and periods of automorphic forms.

Notation and conventions. We fix an algebraic closure $\mathbb{C}$ of the real numbers $\mathbb{R}$ and a choice of $i = \sqrt{-1}$, and we let $\overline{\mathbb{Q}}$ denote the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. We let $c \in \mathrm{Gal}(\mathbb{C}/\mathbb{R})$ denote complex conjugation on $\mathbb{C}$, and we use the same letter to denote its restriction to $\overline{\mathbb{Q}}$. Sometimes we also write $c(z) = \bar{z}$ for $z \in \mathbb{C}$.

Let k be a field. By a variety over k we will mean a geometrically reduced scheme of finite type over k. We let $\mathbb{G}_{\mathrm{m},k}$ denote the usual multiplicative group over k. For any algebraic group G over k, we let $\mathrm{Lie}(G)$ denote its Lie algebra. For us, a reductive group will always be connected. If G is reductive, we let G^{ad} (resp. G^{der}) denote its adjoint group $G/Z(G)$ (resp. its derived subgroup), where $Z(G)$ is the center of G. We let $G^{\mathrm{ab}} = G/G^{\mathrm{der}}$ (a torus). If $T \subset G$ is a torus, we denote by T^{ad} the image of T under the projection $G \to G^{\mathrm{ad}}$. For any commutative group scheme G, we denote by $\mathrm{inv}_G : G \to G$ the map $g \mapsto g^{-1}$.

We denote by $\mathbb{A}$ (resp. $\mathbb{A}_f$) the ring of adèles of $\mathbb{Q}$ (resp. finite adèles). A CM field K is a totally imaginary quadratic extension of a totally real field F.

We let $\mathbb{S} = R_{\mathbb{C}/\mathbb{R}}\mathbb{G}_{\mathrm{m},\mathbb{C}}$. We denote by $c = c_{\mathbb{S}}$ the algebraic automorphism of $\mathbb{S}$ induced by complex conjugation. For any $\mathbb{R}$-algebra A, this is $c \otimes_{\mathbb{R}} \mathrm{id}_A : (\mathbb{C} \otimes_{\mathbb{R}} A)^\times \to (\mathbb{C} \otimes_{\mathbb{R}} A)^\times$ on the points of $\mathbb{S}(A)$. This is often denoted by $z \mapsto \bar{z}$, and on complex points it should not be confused with the other complex conjugation $\mathrm{id}_{\mathbb{C}} \otimes c$ on $\mathbb{S}(\mathbb{C}) = (\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C})^\times$ on the second coordinate.

An involution of a group is an automorphism of order 2, whereas an involution of a ring is an antiautomorphism of order 2. This should not cause any confusion.

We will denote by $\mathbb{H}$ the nonsplit quaternion algebra over $\mathbb{R}$, identified with the set of matrices of the form

$$\begin{pmatrix} x & y \\ -\bar{y} & \bar{x} \end{pmatrix}$$

in $M_2(\mathbb{C})$.

2. Shimura varieties, conjugation, and descent

We will first review some basic facts about Shimura varieties and conjugation by an automorphism of $\mathbb{C}$, specializing to the case of complex conjugation. Then we set up our descent problem, describe some general considerations about reflex fields and Dynkin diagrams, and explain how to construct descent data based on involutions of a Shimura datum.

2.1. ***Shimura varieties.*** A Shimura datum (G, X) will be understood in the sense of Deligne's axioms [1979, (2.1.1.1–3)]. We will assume moreover that the connected component Z^0 of the center Z of G splits over a CM field. For a compact open subgroup $K \subset G(\mathbb{A}_f)$, we put $\mathrm{Sh}_K(G, X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$. For K sufficiently small (which we assume from now on), this complex analytic space is smooth and is equal to the complex points of a complex quasiprojective variety $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. Let $E = E(G, X) \subset \mathbb{C}$ be the reflex field of (G, X); under our hypotheses, this is contained in a CM field, and thus, it is either a CM field or a totally real field. In any case, we let E^+ be the maximal totally real subfield of E. The variety $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ admits a canonical model over E, denoted by $\mathrm{Sh}_K(G, X)_E$. We use the same notation for the pro-objects $\mathrm{Sh}(G, X)(\mathbb{C})$, $\mathrm{Sh}(G, X)_{\mathbb{C}}$, and $\mathrm{Sh}(G, X)_E$. We denote by $w_X : \mathbb{G}_{\mathrm{m},\mathbb{R}} \to G_{\mathbb{R}}$ the composition of $x \in X$ with the weight morphism $w : \mathbb{G}_{\mathrm{m},\mathbb{R}} \to \mathbb{S}$, for some (or any) $x \in X$, and call it the *weight morphism* of (G, X). For $x \in X$, we let $\mu_x : \mathbb{G}_{\mathrm{m},\mathbb{C}} \to G_{\mathbb{C}}$ be the map given by $\mu_x(z) = x_{\mathbb{C}}(z, 1)$, under the identification of $\mathbb{S}_{\mathbb{C}} \cong \mathbb{G}_{\mathrm{m},\mathbb{C}} \times \mathbb{G}_{\mathrm{m},\mathbb{C}}$ given by $(z \otimes a) \mapsto (za, \bar{z}a)$.

We will fix the following notation once and for all. Let $p : G \to G^{\mathrm{ad}}$ be the projection onto G^{ad}. The natural isogeny $Z^0 \times G^{\mathrm{der}} \to G$ and the projection $G \to G^{\mathrm{ab}}$ define an isogeny $Z^0 \to G^{\mathrm{ab}}$. Let $G_1, \dots, G_r$ be the almost simple factors of G^{der} over $\mathbb{Q}$, and let $\widetilde{G}_i \to G_i$ be their simply connected covers. We can write $\widetilde{G}_i = \mathrm{Res}_{F_i/\mathbb{Q}} H_i$, where the fields F_i are totally real and the groups H_i are simply connected, absolutely almost simple over F_i. For each embedding $v \in I_i = \mathrm{Hom}(F_i, \mathbb{C})$, we have groups $H_{i,v} = H_i \otimes_{F_i,v} \mathbb{R}$, and for a fixed $i = 1, \dots, r$, all these groups have the same Dynkin type D_i, which will be called the Dynkin type of $\widetilde{G}_i$ (or of G_i or H_i). We let $I_{i,c} = \{v \in I_i : H^{\mathrm{ad}}_{i,v}(\mathbb{R}) \text{ is compact}\}$ and we let $I_{i,nc}$ be its complement in I_i, which must be nonempty if H_i is nontrivial. We also have that G^{ad} is the direct product of the $G^{\mathrm{ad}}_i = \mathrm{Res}_{F_i/\mathbb{Q}} H^{\mathrm{ad}}_i$, and $G^{\mathrm{ad}}_{\mathbb{R}}$ is the direct product of the $H^{\mathrm{ad}}_{i,v}$ for $i = 1, \dots, r$ and $v \in I_i$. Let X^{ad} be the $G^{\mathrm{ad}}(\mathbb{R})$-conjugacy class containing $p_{\mathbb{R}}(X)$, and write $X^{\mathrm{ad}} = \prod_{i,v} X_{i,v}$ with $X_{i,v}$ an $H^{\mathrm{ad}}_{i,v}(\mathbb{R})$-conjugacy class of morphisms $\mathbb{S} \to H^{\mathrm{ad}}_{i,v}$. For each i and each $v \in I_{i,nc}$, there is a special node $s_{i,v}$ in the Dynkin diagram $D_{i,v}$ of $H_{i,v}$ attached to $X_{i,v}$, which uniquely determines $X_{i,v}$ as a conjugacy class with target $H^{\mathrm{ad}}_{i,v}$ (in the sense that if Y is an $H^{\mathrm{ad}}_{i,v}(\mathbb{R})$-conjugacy class satisfying Deligne's axioms, for which its associated special node is $s_{i,v}$, then $Y = X_{i,v}$ [Deligne 1979, §1.2.6]).

2.2. ***Conjugation.*** For the general properties of conjugation of Shimura varieties, we mainly follow [Milne 1990; Milne and Shih 1982b; Deligne 1982; Milne and Shih 1982a]; see also [Langlands 1979]. Let (G, X) be a Shimura datum. A special pair (T, x) consists of a maximal torus $T \subset G$ and a point $x \in X$ factoring through $T_{\mathbb{R}}$. Fix $x \in X$ a special point, and let $\sigma \in \mathrm{Aut}(\mathbb{C})$. We denote by $({}^{\sigma,x}G, {}^{\sigma,x}X)$ the conjugate Shimura datum. We recall its construction below. By Theorem II.4.2 of

[Milne 1990], there exists a unique isomorphism

$$\varphi_{\sigma,x} : \sigma\,\mathrm{Sh}(G, X)_{\mathbb{C}} = \mathrm{Sh}(G, X)_{\mathbb{C}} \times_{\mathbb{C},\sigma} \mathbb{C} \simeq \mathrm{Sh}({}^{\sigma,x}G, {}^{\sigma,x}X)_{\mathbb{C}}$$

satisfying certain conditions. Choosing a different special point gives canonically isomorphic results [Milne 1990, Proposition II.4.3]. The reflex field of $({}^{\sigma,x}G, {}^{\sigma,x}G)$ is $\sigma(E)$, and $\varphi_{\sigma,x}$ identifies $\sigma\,\mathrm{Sh}(G, X)_E = \mathrm{Sh}(G, X)_E \times_{E,\sigma} \sigma(E)$ with the canonical model of $\mathrm{Sh}({}^{\sigma,x}G, {}^{\sigma,x}G)_{\mathbb{C}}$ over $\sigma(E)$ [Milne 1990, Theorem II.5.5]. In particular, if $\sigma(E) = E$, then $\varphi_{\sigma,x}$ defines an isomorphism

$$\varphi_{\sigma,x} : \sigma\,\mathrm{Sh}(G, X)_E \simeq \mathrm{Sh}({}^{\sigma,x}G, {}^{\sigma,x}X)_E$$

over E. All of this also works at finite level: if $K \subset G(\mathbb{A}_f)$ is compact open, $\varphi_{\sigma,x}$ sends $\sigma\,\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ to $\mathrm{Sh}_{{}^{\sigma,x}K}({}^{\sigma,x}G, {}^{\sigma,x}X)_{\mathbb{C}}$ (same thing replacing $\mathbb{C}$ by E and $\sigma(E)$), where ${}^{\sigma,x}K \subset {}^{\sigma,x}G(\mathbb{A}_f)$ is explicit (see below).

We are interested mainly in the case $\sigma = c$, but nevertheless it will be useful to recall the general construction of $({}^{\sigma,x}G, {}^{\sigma,x}X)$. Let $\mathfrak{S}$ be the (connected) Serre group. This can be defined as the group of automorphisms of the forgetful fiber functor from the Tannakian category of CM $\mathbb{Q}$-Hodge structures to the category of finite-dimensional $\mathbb{Q}$-vector spaces. (Here a $\mathbb{Q}$-Hodge structure is a $\mathbb{Q}$-vector space V such that $V \otimes \mathbb{C}$ is endowed with a Hodge structure; the structure is CM if the algebra of elements of $\mathrm{End}(V)$ which induce morphisms of Hodge structure contains a commutative semisimple subalgebra of dimension $\dim_{\mathbb{Q}}(V)$.) Let $\mathfrak{T}$ denote the Taniyama group, defined here as the group of automorphisms of the Betti fiber functor in Deligne's Tannakian category of CM motives for absolute Hodge cycles over $\mathbb{Q}$; this is the Tannakian category generated by Artin motives and by the cohomology of abelian varieties over $\mathbb{Q}$ which are potentially CM. These are proalgebraic groups, and there is a natural exact sequence

$$1 \to \mathfrak{S} \to \mathfrak{T} \xrightarrow{\pi} \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to 1,$$

where the second arrow corresponds to the functor taking a CM motive M to its CM Hodge structure $H_B(M)$, and π corresponds to the natural inclusion of the category of Artin motives into the category of CM motives. The group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is to be considered as the proalgebraic group given by the inverse limit of the finite constant groups $\mathrm{Gal}(L/\mathbb{Q})$, for $L \subset \mathbb{C}$ a finite Galois extension of $\mathbb{Q}$. There is a continuous section of π over $\mathbb{A}_f$ denoted by

$$\mathrm{sp} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathfrak{T}(\mathbb{A}_f).$$

For a motive M, $\mathrm{sp}(\sigma)$ corresponds to the automorphism of $H_B(M) \otimes_{\mathbb{Q}} \mathbb{A}_f$ obtained from the Galois action of σ on étale cohomology using the comparison isomorphism.

Suppose that $\sigma = c$ is complex conjugation. Then $\mathrm{sp}(c) \in \mathfrak{T}(\mathbb{A}_f)$ can be described as follows, as explained in [Deligne 1982, Lemme 5]. Suppose that M is a CM motive over $\mathbb{Q}$, realized as the cohomology of an algebraic variety X over $\mathbb{Q}$. The action of complex conjugation on $X(\mathbb{C})$ induces an involution F on the Betti realization $H_B(M) = H^i(X(\mathbb{C}), \mathbb{Q})$. The automorphism $\mathrm{sp}(c) : H_B(M) \otimes \mathbb{A}_f \to H_B(M) \otimes \mathbb{A}_f$ is then equal to $F \otimes_{\mathbb{Q}} \mathrm{id}_{\mathbb{A}_f}$. This implies, in particular, that $\mathrm{sp}(c) \in \mathfrak{T}(\mathbb{Q})$.

For any $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, we let ${}^{\sigma}\mathfrak{S} = \pi^{-1}(\sigma)$. There is a cocharacter $\mu_{\mathrm{can}} : \mathbb{G}_{\mathrm{m},\mathbb{C}} \to \mathfrak{S}_{\mathbb{C}}$, which in Tannakian terms gives rise to the Hodge cocharacter of the Hodge structures on $H_B(M) \otimes_{\mathbb{Q}} \mathbb{C}$.

Let G be any algebraic group over $\mathbb{Q}$ and $\rho : \mathfrak{S} \to G^{\mathrm{ad}}$ be a homomorphism, inducing an action of $\mathfrak{S}$ on G by group automorphisms (conjugation). Let ${}^{\sigma,\rho}G = {}^{\sigma}\mathfrak{S} \times_{\mathfrak{S},\rho} G$ be the group obtained by twisting G by the torsor ${}^{\sigma}\mathfrak{S}$. Thus, ${}^{\sigma,\rho}G$ is the fpqc sheaf associated with the presheaf sending a $\mathbb{Q}$-algebra R to the group ${}^{\sigma}\mathfrak{S}(R) \times_{\mathfrak{S}(R),\rho} G(R)$, which is the quotient of ${}^{\sigma}\mathfrak{S}(R) \times G(R)$ by the right action $(s, g)s_1 = (ss_1, s_1^{-1}g)$ of $\mathfrak{S}(R)$. The class of an element (s, g) in this quotient will be denoted by $s \cdot g$.

Lemma 2.2.1. *Keep the notation and assumptions as above, with $\sigma = c$. There exists a natural isomorphism ${}^{c,\rho}G \to G$.*

Proof. As explained above, $\mathrm{sp}(c) \in {}^{c}\mathfrak{S}(\mathbb{Q})$, so ${}^{c}\mathfrak{S}$ is trivialized over $\mathbb{Q}$. In particular, the map $\mathrm{sp}(c)_R \cdot g \mapsto g$ (for $g \in G(R)$) defines a group isomorphism between the presheaves defining ${}^{c,\rho}G$ and G. A fortiori, this defines an isomorphism ${}^{c,\rho}G \to G$. □

Remark 2.2.2. If $H \subset G$ is a subgroup on which $\mathfrak{S}$ acts trivially, then ${}^{c,\rho}H$ is canonically isomorphic to H (this is true for any σ). This identification is compatible with that of Lemma 2.2.1.

Remark 2.2.3. In [Milne 1990, §II.4], an isomorphism $G(\mathbb{A}_f) \to {}^{c,\rho}G(\mathbb{A}_f)$ is constructed, which is denoted by $g \mapsto {}^{c}g$. When identifying ${}^{c,\rho}G$ with G using Lemma 2.2.1, this becomes the identity map. A similar remark applies to the isomorphism $g \mapsto {}^{c}g$ between $G_{\mathbb{C}}$ and ${}^{c,\rho}G_{\mathbb{C}}$ defined in [Milne 1990, §III.1] (note that the element $z_\infty(c)$ defined in [op. cit.] is equal to $\mathrm{sp}(c)_{\mathbb{C}}$).

Suppose that (G, X) is a Shimura datum as before, and (T, x) is a special pair. The map μ_x factors through $T_{\mathbb{C}}$, and there exists a unique homomorphism

$$\rho_x^{\mathrm{ad}} : \mathfrak{S} \to G^{\mathrm{ad}}$$

such that $(\rho_x^{\mathrm{ad}})_{\mathbb{C}} \circ \mu_{\mathrm{can}} = \mu_x^{\mathrm{ad}}$. For $\sigma \in \mathrm{Aut}(\mathbb{C})$, the group ${}^{\sigma,x}G$ is defined to be ${}^{\sigma,\rho_x^{\mathrm{ad}}}G$ in the previous notation (where we take the restriction of σ to $\overline{\mathbb{Q}}$). Since the cocharacter $\sigma(\mu_x)$ of $T = {}^{\sigma,\rho_x}T$ commutes with its complex conjugate, it is the

Hodge cocharacter associated with a map $\mathbb{S} \to {}^{\sigma,x}G_{\mathbb{R}}$ which we denote by

$$^{\sigma}x : \mathbb{S} \to {}^{\sigma,x}G_{\mathbb{R}}.$$

Finally, ${}^{\sigma,x}X$ is defined to be the ${}^{\sigma,x}G(\mathbb{R})$-conjugacy class of ${}^{\sigma}x$.

Assume now that $\sigma = c$. By Lemma 2.2.1, we can identify ${}^{c,x}G$ with G, and hence, we can see ${}^{c}x : \mathbb{S} \to {}^{c,x}G_{\mathbb{R}}$ as a map

$$^{c}x : \mathbb{S} \to G_{\mathbb{R}}$$

with target $G_{\mathbb{R}}$. Let $c = c_{\mathbb{S}} : \mathbb{S} \to \mathbb{S}$ be complex conjugation on $\mathbb{S}$. For any $h : \mathbb{S} \to G_{\mathbb{R}}$, let $\bar{h} = h \circ c$.

Lemma 2.2.4. *In the notation above, we have that*

$$^{c}x = \bar{x}.$$

Proof. It is enough to show that $\bar{x}(\mathbb{C}) = ({}^{c}x)(\mathbb{C}) : \mathbb{S}(\mathbb{C}) \to G(\mathbb{C})$. Recall that we are identifying $\mathbb{S}(\mathbb{C}) = (\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C})^{\times}$ with $\mathbb{C}^{\times} \times \mathbb{C}^{\times}$ via the map $(z \otimes a) \mapsto (za, \bar{z}a)$. Then $c = c_{\mathbb{S}} : \mathbb{S}(\mathbb{C}) \to \mathbb{S}(\mathbb{C})$, which is given by $c(z \otimes a) = \bar{z} \otimes a$, becomes the map $(a, b) \mapsto (b, a)$. This is an algebraic automorphism of $\mathbb{S}$. There is another complex conjugation, which will be denoted by c' here, on the complex points of $\mathbb{S}$. Namely,

$$c' : \mathbb{S}(\mathbb{C}) \to \mathbb{S}(\mathbb{C}),$$

which is induced by complex conjugation on $\mathbb{C}$. It is given by $c'(z \otimes a) = z \otimes \bar{a}$. Then, as a map on $\mathbb{C}^{\times} \times \mathbb{C}^{\times}$, it is given by $(a, b) \mapsto (\bar{b}, \bar{a})$.

Recall that $\mu_x(a) = x(\mathbb{C})(a, 1)$ for $a \in \mathbb{C}^{\times}$. For readability purposes, we use the notation $x_{\mathbb{C}}(a, b)$ instead of $x(\mathbb{C})(a, b)$ in what follows. Then, for $a, b \in \mathbb{C}^{\times}$, we have that

$$\bar{x}_{\mathbb{C}}(a, b) = x_{\mathbb{C}}(b, a) = x_{\mathbb{C}}(b, 1)x_{\mathbb{C}}(1, a). \tag{2.2.5}$$

Similarly,

$$({}^{c}x)_{\mathbb{C}}(a, b) = \mu_{{}^{c}x}(a)\mu_{\overline{{}^{c}x}}(b). \tag{2.2.6}$$

Now, by definition, $\mu_{{}^{c}x} = c(\mu_x)$, where c now denotes the action on cocharacters. If we let $g \mapsto \bar{g}$ denote the map on $G(\mathbb{C})$ induced by $c : \mathbb{C} \to \mathbb{C}$, then

$$\mu_{{}^{c}x}(a) = \overline{\mu_x(\bar{a})} = \overline{x_{\mathbb{C}}(\bar{a}, 1)} \tag{2.2.7}$$

for $a \in \mathbb{C}^{\times}$. Since x is defined over $\mathbb{R}$, it commutes with the maps on complex points induced by $c : \mathbb{C} \to \mathbb{C}$. That is, the diagram

$$\begin{array}{ccc} \mathbb{S}(\mathbb{C}) & \xrightarrow{x_{\mathbb{C}}} & G(\mathbb{C}) \\ \downarrow{\scriptstyle c'} & & \downarrow{\scriptstyle g \mapsto \bar{g}} \\ \mathbb{S}(\mathbb{C}) & \xrightarrow{x_{\mathbb{C}}} & G(\mathbb{C}) \end{array}$$

is commutative. From this and (2.2.7), it follows that

$$\mu_{^c x}(a) = x_{\mathbb{C}}(1, a). \tag{2.2.8}$$

Similarly,

$$\mu_{\overline{^c x}}(b) = \overline{^c x}_{\mathbb{C}}(b, 1) = ({}^c x)_{\mathbb{C}}(1, b) = \mu_{^c({}^c x)}(b).$$

But $\mu_{^c({}^c x)} = c(\mu_{^c x}) = \mu_x$, so this equals $x_{\mathbb{C}}(b, 1)$. The proof finishes by combining (2.2.5), (2.2.6), and (2.2.8). □

By the last lemma, we can identify ${}^{c,x}X$ with the $G(\mathbb{R})$-conjugacy class of $\bar{x}$. The map $h \mapsto \bar{h}$ defines an antiholomorphic isomorphism between X and ${}^{c,x}X$. This does not depend on x, and from now on we let

$$\overline{X} = \{\bar{h} : h \in X\}.$$

Thus, the pair $({}^{c,x}G, {}^{c,x}X)$ becomes naturally identified with the pair $(G, \overline{X})$. The isomorphism $\varphi_{c,x}$ becomes, under this identification, an isomorphism $\varphi : \mathrm{Sh}_K(G, X)_E \times_{E,c} E \to \mathrm{Sh}_K(G, \overline{X})_E$. On complex points, it defines an antiholomorphic isomorphism between $\mathrm{Sh}_K(G, X)(\mathbb{C})$ and $\mathrm{Sh}_K(G, \overline{X})(\mathbb{C})$, which we denote by ϕ. For $[h, g] \in \mathrm{Sh}_K(G, X)(\mathbb{C})$, we have that $\phi([h, g]) = [\bar{h}, g] \in \mathrm{Sh}_K(G, \overline{X})(\mathbb{C})$.

For example, suppose that $E \subset \mathbb{R}$. Then there is an antiholomorphic involution on $\mathrm{Sh}_K(G, X)(\mathbb{C})$ defined by complex conjugation acting on $\mathbb{C}$. It follows from the theory of canonical models that this involution takes the form $[h, g] \mapsto [\eta(h), g]$, where $\eta : X \to X$ is an antiholomorphic involution of the form $\eta(g \cdot x) = (gn) \cdot x$ for some $n \in N(\mathbb{R})$ (here N is the normalizer in G of T). See [Milne 1990, §II.7] for details. In fact, the theory implies that there exists $n \in N(\mathbb{R})$ such that ${}^c x = n \cdot x$, and thus, $\overline{X} = X$. Then the map η becomes what we called ϕ; that is, $\eta(h) = \bar{h}$ for any $h \in X$.

2.3. ***Involutions of Shimura data and descent.*** Fix a Shimura datum (G, X), with reflex field E. For an involution $\theta : G \to G$, let $\theta(X)$ be the $G(\mathbb{R})$-conjugacy class $\{\theta(h) : h \in X\}$, where $\theta(h) = \theta_{\mathbb{R}} \circ h$. Since we want to consider involutions θ that send X to $\overline{X} \neq X$, from now on, we will focus on the case where E is a CM field (if E is totally real, the identity map on G takes X to $\overline{X}$). Let $E^+ \subset E$ be the maximal totally real subfield, and let $\iota \in \mathrm{Gal}(E/E^+)$ be the nontrivial automorphism, i.e., the restriction of complex conjugation c to E.

Suppose that θ is an involution of G such that $\theta(X) = \overline{X}$. For a compact open subgroup $K \subset G(\mathbb{A}_f)$, denote by ${}^\theta K = \theta(K) \subset G(\mathbb{A}_f)$. Then θ induces an isomorphism of algebraic varieties $\mathrm{Sh}(\theta) : \mathrm{Sh}_K(G, X)_E \to \mathrm{Sh}_{^\theta K}(G, \overline{X})_E$. On complex points, this takes $[h, g]$ to $[\theta_{\mathbb{R}} \circ h, \theta(g)]$. Suppose that ${}^\theta K = K$. Then $\mathrm{Sh}(\theta)^{-1} \circ \varphi$ defines an isomorphism $\psi : \iota(\mathrm{Sh}_K(G, X)_E) = \mathrm{Sh}_K(G, X)_E \times_{E,\iota} E \to \mathrm{Sh}_K(G, X)_E$.

Let V be an arbitrary scheme over E. Recall that an E/E^+-descent datum is a pair of isomorphisms $\psi_{\mathrm{id}} : \mathrm{id}(V) = V \times_{E,\mathrm{id}} E \to V$ and $\psi_\iota : \iota V = V \times_{E,\iota} E \to V$ of schemes over E satisfying the cocycle condition

$$\psi_\sigma \circ \sigma(\psi_\tau) = \psi_{\sigma\tau}$$

for all $\sigma, \tau \in \mathrm{Gal}(E/E^+)$, using the natural identification $\sigma(\tau(V)) = (\sigma\tau)V$. Then necessarily ψ_{id} is the first projection $\mathrm{id}(V) \to V$, and thus, to give a descent datum amounts to give an isomorphism $\psi = \psi_\iota : \iota(V) \to V$ such that $\psi \circ \iota(\psi) : \iota(\iota(V)) \to V$ is equal to the identity map, when identifying $\iota(\iota(V)) = V$. By definition, such a descent datum is effective if there exists a scheme V_0 over E^+ and an isomorphism $m : V \to V_{0,E} = V_0 \times_{E^+} E$ such that $m \circ \psi = \iota(m)$, after identifying $\iota(V_{0,E}) = V_{0,E}$. If V is a quasiprojective algebraic variety, then any descent datum for V is effective. This was first proved by Weil [1956]. For a modern reference, see [Bosch et al. 1990, §6.2].

Theorem 2.3.1. *The map* $\psi : \iota(\mathrm{Sh}_K(G, X)_E) \to \mathrm{Sh}_K(G, X)_E$ *obtained as above from an involution* $\theta : G \to G$ *such that* $\theta(X) = \overline{X}$ *and* ${}^{\theta}K = K$ *is an effective* E/E^+*-descent datum on the Shimura variety* $\mathrm{Sh}_K(G, X)_E$*. Hence, there exists a quasiprojective, smooth, algebraic variety* $\mathrm{Sh}_K(G, X)_{E^+}$ *over* E^+*, and an isomorphism* $m : \mathrm{Sh}_K(G, X)_E \to \mathrm{Sh}_K(G, X)_{E^+} \times_{E^+} E$ *such that* $m \circ \psi = \iota(m)$*.*

Proof. Let $V = \mathrm{Sh}_K(G, X)_E$ and $\overline{V} = \mathrm{Sh}_K(G, \overline{X})_E$, and let $n : V \to \iota(\iota(V))$ be the natural isomorphism. We need to check that $\psi \circ \iota(\psi) \circ n = \mathrm{id}_V$, and for this it is enough to see that both morphisms are equal on the set of complex points $V(\mathbb{C})$. Let $c_V : V(\mathbb{C}) \to (\iota V)(\mathbb{C})$ be the bijection that sends $x : \mathrm{Spec}(\mathbb{C}) \to V$ to $p_{\iota,V}^{-1} \circ x \circ \mathrm{Spec}(c)$, where $p_{\iota,V} : \iota V \to V$ is the first projection, and define $c_{\iota V} : (\iota V)(\mathbb{C}) \to (\iota(\iota V))(\mathbb{C})$ similarly. Then we have that $n(\mathbb{C}) = c_{\iota V} \circ c_V$, $\iota(\psi)(\mathbb{C}) = c_V \circ \psi(\mathbb{C}) \circ c_{\iota V}^{-1}$, and ψ satisfies that $\psi(\mathbb{C}) \circ c_V = \mathrm{Sh}(\theta)^{-1}(\mathbb{C}) \circ \phi$. Recall that $\phi : V(\mathbb{C}) \to \overline{V}(\mathbb{C})$ sends $[h, g]$ to $[\bar{h}, g]$. Putting all this together, we get that

$$(\psi \circ \iota(\psi) \circ n)(\mathbb{C}) = \mathrm{Sh}(\theta)^{-1}(\mathbb{C}) \circ \phi \circ \mathrm{Sh}(\theta)^{-1}(\mathbb{C}) \circ \phi,$$

and thus,

$$(\psi \circ \iota(\psi) \circ n)(\mathbb{C})([h, g]) = \big[\theta^{-1}\big(\overline{\theta^{-1}(\bar{h})}\big), \theta^{-2}(g)\big].$$

But for any $y \in \overline{X}$, $\theta^{-1}(y) = \theta_{\mathbb{R}}^{-1} \circ y$, and so

$$\theta^{-1}\big(\overline{\theta^{-1}(\bar{h})}\big) = \theta^{-1}\big(\overline{\theta_{\mathbb{R}}^{-1} \circ \bar{h}}\big) = \theta^{-1}\big(\overline{\theta_{\mathbb{R}}^{-1} \circ h \circ c}\big) = \theta^{-1}(\theta_{\mathbb{R}}^{-1} \circ h) = \theta_{\mathbb{R}}^{-2} \circ h = \theta^{-2}(h),$$

and thus, $(\psi \circ \iota(\psi) \circ n)(\mathbb{C})([h, g]) = [h, g]$, using the fact that $\theta^2 = \mathrm{id}$. Finally, since $\mathrm{Sh}_K(G, X)_E$ is quasiprojective, the descent datum just constructed is effective. □

Remark 2.3.2. The model of Theorem 2.3.1 depends on the descent datum, which in turns depends on the particular involution θ.

We note that, by the nature of the descent datum, Hecke operators do not descend to the model $\mathrm{Sh}_K(G, X)_{E^+}$. Given $q \in G(\mathbb{A}_f)$, the Hecke operator T_q is a morphism of algebraic varieties $T_q : \mathrm{Sh}_K(G, X)_E \to \mathrm{Sh}_{q^{-1}Kq}(G, X)_E$, which on complex points is given by $T_q([h, g]) = [h, gq]$. Then $T_{\theta(q)} \circ \mathrm{Sh}(\theta) = \mathrm{Sh}(\theta) \circ T_q : \mathrm{Sh}_K(G, X)_E \to \mathrm{Sh}_{\theta(q)^{-1}\theta K\theta(q)}(G, \overline{X})_E$. The Hecke operator T_q descends to a map $\mathrm{Sh}_K(G, X)_{E^+} \to \mathrm{Sh}_{q^{-1}Kq}(G, X)_{E^+}$ if and only if $T_{\theta(q)} = T_q$.

In the following sections we will construct several examples of involutions θ as above, and explain a general framework for such constructions.

3. Opposition involutions

In this section we recall some basic facts about opposition involutions and prove a few results that will be needed in the forthcoming sections. For the basic facts regarding root data, see [Springer 1979].

3.1. ***Root data.*** Let $\Psi = (X, \Phi, X^\vee, \Phi^\vee)$ be a root datum with $\Phi \neq \varnothing$. Let Q be the subgroup of X generated by Φ, and $V = Q \otimes_Z \mathbb{Q}$. Let $W = W(\Phi)$ be the Weyl group of the root system Φ in V. This can be naturally identified with the Weyl group of $\Phi^\vee$ and with the subgroup of $\mathrm{Aut}_{\mathbb{Z}}(X)$ generated by the reflections s_α for $\alpha \in \Phi$. Choose a basis Δ, and consider the associated based root datum $\Psi_0 = (X, \Phi, \Delta, X^\vee, \Phi^\vee, \Delta^\vee)$.

There is an obvious notion of isomorphism of root data (resp. based root data) $\Psi \to \Psi'$ (resp. $\Psi_0 \to \Psi'_0$). It amounts to giving a $\mathbb{Z}$-linear isomorphism $f : X \to X'$ such that $f(\Phi) = \Phi'$ and ${}^t f(f(\alpha)^\vee) = \alpha^\vee$ for all $\alpha \in \Phi$ (resp. and $f(\Delta) = \Delta'$). Here ${}^t f$ denotes the transpose with respect to the root data pairings. We denote by $\mathrm{Aut}(\Psi)$ (resp. $\mathrm{Aut}(\Psi_0)$) the group of automorphisms of Ψ (resp. Ψ_0). Each s_α can be seen as an automorphism of Ψ, and thus, there is a natural inclusion $W \subset \mathrm{Aut}(\Psi)$. We also denote by $-1 \in \mathrm{Aut}(\Psi)$ the automorphism that sends $x \in X$ to $-x \in X$.

Assume from now on that Φ is reduced. If Δ is a basis, let w_0 be the longest element of W with respect to it. Then $w_0(\Delta) = -\Delta$, and thus, $-w_0 = -1 \circ w_0 \in \mathrm{Aut}(\Psi_0)$. We call $\star = -w_0$ the *opposition involution* of Ψ_0 (since $w_0^2 = 1$ it is indeed an involution). We denote the action of $\star$ on elements x (which can be characters of T, nodes of the Dynkin diagram, etc.) by $x \mapsto x^\star$. When $\Phi = \varnothing$, in which case Ψ is called toral, we directly define $\star = -1 \in \mathrm{Aut}_{\mathbb{Z}}(X)$.

Remark 3.1.1. An isogeny (in particular, an isomorphism) of based root data will commute with the corresponding opposition involutions. In particular, $\star$ is a central element of $\mathrm{Aut}\,\Psi_0$.

Remark 3.1.2. Let $X_0 \subset X$ denote the subgroup of X orthogonal to $\Phi^\vee$. The root datum Ψ is called semisimple when $X_0 = 0$. If this is not the case, then there exists a nonzero $x \in X_0$, which hence must be invariant under W. In particular,

$x^\star = -x \neq x$, so $\star$ cannot be the identity map if the root datum is not semisimple. In the same vein, if the root datum is toral, then $\star \neq 1$ unless Ψ is trivial (that is, also semisimple).

Suppose now that k is an algebraically closed field of characteristic 0, and let G be a reductive group over k. Let $T \subset G$ be a maximal torus, and $\Psi = \Psi(G, T)$ be the associated root datum, so that $X = X^*(T)$. Let $B \supset T$ be a Borel subgroup, and let $\Psi_0 = \Psi_0(G, T, B)$ be the corresponding based root datum. Let $\mathrm{Aut}(G)$ be the group of automorphisms of G, and $\mathrm{Inn}(G) \subset \mathrm{Aut}(G)$ be the subgroup of inner automorphisms (that is, defined by elements in $G(k)$). Thus, $\mathrm{Inn}(G) \simeq G^{\mathrm{ad}}(k) \simeq G(k)/Z(k)$, where Z is the center of G. Then there is a split exact sequence

$$1 \to \mathrm{Inn}(G) \to \mathrm{Aut}(G) \to \mathrm{Aut}\,\Psi_0 \to 1 \tag{3.1.3}$$

where, for $f \in \mathrm{Aut}(G)$, the third arrow sends f to the automorphism of Ψ_0 induced by $f' \in \mathrm{Aut}(G, T, B)$, where $f' = \mathrm{int}(g) \circ f$ for any element $g \in G(k)$ such that $\mathrm{int}(g) f(B, T) = (B, T)$. We define an *opposition involution* of G (with respect to (B, T)) to be any element $\theta \in \mathrm{Aut}(G)$ of order 1 or 2 that induces the opposition involution $\star$ in $\mathrm{Aut}\,\Psi_0$. Note that this definition does not require θ to preserve T or B. If θ' is another such involution, then $\theta' = \mathrm{int}(g) \circ \theta$ for some $g \in G(k)$. If θ is an opposition involution for (B, T) and (B', T') is another Borel pair, then it is also an opposition involution for (B', T'). The exact sequence (3.1.3) is split by the choice of a pinning. More precisely, let $\Delta \subset \Phi$ be the set of simple roots corresponding to B. For each $\alpha \in \Delta$, let $U_\alpha \in G$ be the root group of α [Springer 1979, §2.3], and let $u_\alpha \in U_\alpha$ be a nontrivial element. The *pinning* is the datum $\{u_\alpha\}_{\alpha \in \Delta}$ with respect to (B, T), and a splitting $\mathrm{Aut}\,\Psi_0 \to \mathrm{Aut}(G)$ of (3.1.3) associated with this pinning is given by an isomorphism $\mathrm{Aut}\,\Psi_0 \simeq \mathrm{Aut}(G, T, B, \{u_\alpha\}_{\alpha\in\Delta})$; two such splittings differ by an automorphism $\mathrm{int}(t)$ for some $t \in T(k)$. In particular, after choosing a pinning, we can take $\theta \in \mathrm{Aut}(G)$ to be the image of $\star$ under the splitting and this will be an opposition involution, which proves their existence. Note that we are actually showing that there are opposition involutions in $\mathrm{Aut}(G)$ which preserve T and B (and a fixed pinning).

Let k be any field of characteristic 0, and $\bar{k}$ be an algebraic closure of k. Let $\Gamma = \mathrm{Aut}(\bar{k}/k)$. Let G be a reductive group over k, $T \subset G$ a maximal torus, and $B \supset T_{\bar{k}}$ a Borel subgroup of $G_{\bar{k}}$. Let $\Psi = \Psi(G_{\bar{k}}, T_{\bar{k}})$ and $\Psi_0 = \Psi_0(G_{\bar{k}}, T_{\bar{k}}, B)$. There is a natural action of Γ on X, denoted by $\chi \mapsto {}^\gamma\chi$, where

$${}^\gamma\chi(t) = \gamma(\chi(\gamma^{-1}(t)))$$

for $\gamma \in \Gamma$ and $t \in T(\bar{k})$. We call it the *usual* action of Γ on X. It defines an action of Γ on Ψ. Let $\gamma \in \Gamma$. Then we define a second action $\mu_G(\gamma)$ on X, the $*$-action, given by $\mu_G(\gamma)(\chi)(t) = {}^\gamma\chi(n^{-1}tn)$ for $t \in T(\bar{k})$, where $n \in G(\bar{k})$ is an element such that $\mathrm{int}(n)$ sends the Borel pair $(\gamma(B), \gamma(T_{\bar{k}}))$ to $(B, T_{\bar{k}})$. For example, if B is a Borel defined over k, then we can take $n = 1$ and the $*$-action is just the usual action $\chi \mapsto {}^{\gamma}\chi$. Going back to the general case, this gives a morphism $\mu_G : \Gamma \to \mathrm{Aut}\,\Psi_0$, and it induces an action of Γ on $\mathrm{Aut}\,\Psi_0$ by taking $\rho \mapsto \mu_G(\gamma) \circ \rho \circ \mu_G(\gamma)^{-1}$ for $\rho \in \mathrm{Aut}\,\Psi_0$. There is also an action of Γ on $\mathrm{Aut}(G_{\bar{k}})$ given by $\gamma \cdot f = (\mathrm{id}_G \times_{\mathrm{Spec}(k)} \mathrm{Spec}(\gamma^{-1})) \circ f \circ (\mathrm{id}_G \times_{\mathrm{Spec}(k)} \mathrm{Spec}(\gamma))$, which on $G(\bar{k})$-points is simply $g \mapsto \gamma(f(\gamma^{-1}(g)))$. It preserves the subgroup $\mathrm{Inn}(G_{\bar{k}}) = G(\bar{k})/Z(\bar{k})$, where it acts as usual. The exact sequence (3.1.3) becomes

$$1 \to \mathrm{Inn}(G_{\bar{k}}) \to \mathrm{Aut}(G_{\bar{k}}) \to \mathrm{Aut}\,\Psi_0 \to 1 \tag{3.1.4}$$

and is Γ-equivariant. We define an opposition involution of G to be an automorphism $\theta \in \mathrm{Aut}(G)$ of order 1 or 2 such that $\theta_{\bar{k}}$ is an opposition involution on $G_{\bar{k}}$.

There may not be a Γ-equivariant splitting of (3.1.4), so it may not always be possible to construct in this way an opposition involution of G. However, if G is quasisplit and B is a Borel subgroup defined over k, it can be shown [Demazure 1965/66, §3.10] that there exists a Γ-equivariant splitting. Since $\star \in \mathrm{Aut}\,\Psi_0$ is central, it commutes with $\mu_G(\gamma)$ for any $\gamma \in \Gamma$, and thus, it is a Γ-invariant element in the last group of (3.1.4). Thus, for quasisplit reductive groups over k, there always exist opposition involutions on G over k, but the condition of G being quasisplit is far from necessary. There are many nonquasisplit cases where the opposition involution is trivial (see below), and so obviously defined over k. There are many nontrivial examples as well, as we will see later.

Remark 3.1.5. If $G = T$ is a torus, then there exists one and only one opposition involution $\theta \in \mathrm{Aut}(G)$, namely $\theta = \mathrm{inv}_G$.

Lemma 3.1.6. *If θ is an opposition involution of G, then $\theta_Z : Z \to Z$ is equal to* inv_Z.

Proof. It is enough to see that both maps induce the same map on $X^*(Z)$, that is, that $\theta_Z^* : X^*(Z) \to X^*(Z)$ is multiplication by -1, and thus, we can assume that $k = \bar{k}$. Let (B, T) be a Borel pair. Then $Z \subset T$. Let $\chi \in X^*(Z)$. Then there exists $\mu \in X^*(T)$ such that $\mu|_Z = \chi$. We claim that $\theta_Z^*(\chi) = (\mu^\star)|_Z$. Indeed, for $z \in Z(k)$, $\theta_Z^*(\chi)(z) = \chi(\theta(z))$, whereas $(\mu^\star)|_Z(z) = \Psi_0(\theta)(\mu)(z) = \mu((\mathrm{int}(g) \circ \theta)(z)) = \mu(\theta(z))$ (where $g \in G(k)$ sends $\theta(B, T)$ to (B, T)), which shows that $\theta_Z^* = (\mu^\star)|_Z$.

On the other hand, if $n_0 \in N_G(T)(k)$ represents $w_0 \in W = N_G(T)(k)/T(k)$, then for $z \in Z(k)$, $\mu^\star(z) = \mu(n_0^{-1} z^{-1} n_0) = \mu(z^{-1}) = \mu^{-1}(z)$ because $z \in Z(k)$. Thus, $\theta_Z^*(\chi) = -\chi$, as desired, where we have switched back to the additive notation for the group $X^*(Z)$. □

Remark 3.1.7. The last lemma shows in particular that if the identity map is an opposition involution, then Z is killed by 2. Then Z^0 must be trivial; that is, G must be semisimple (see also Remark 3.1.2).

3.2. ***Dynkin diagrams and special nodes.*** Let Ψ_0 be a based root datum with $\Phi \neq \varnothing$ and reduced, and let $\mathcal{D}$ be its Dynkin diagram. Then the opposition involution $\star$ acts on $\mathcal{D}$. We include for reference the list of connected Dynkin diagrams and their opposition involutions; see [Bourbaki 2002] for notation of nodes and more details. We also list the special nodes of each diagram (see [Deligne 1979, §1.2.5], for the definition of special node). Also, note that if Ψ is semisimple, then $\star$ is trivial on Ψ_0 if and only if it is trivial on $\mathcal{D}$. For a Shimura datum (G, X), the only factors of G^{ad} that contribute to a CM reflex field are the ones of type A_l ($l \geq 2$), D_l ($l \geq 5$ odd), or E_6. This follows from the list below and Proposition 2.3.6 of [Deligne 1979]:

- $\mathcal{D} = A_l$ ($l \geq 1$).
  $\alpha_i^\star = \alpha_{l+1-i}$ (so $\star$ is trivial if $l = 1$).
  All nodes α_i are special.
- $\mathcal{D} = B_l$ ($l \geq 2$) or C_l ($l \geq 3$).
  $\star$ is trivial.
  There is only one special node: α_1 in the B_l case, and α_l in the C_l case.
- $\mathcal{D} = D_l$ ($l \geq 4$).
  If l is even, $\star$ is trivial.
  If l is odd, $\alpha_i^\star = \alpha_i$ for $i < l - 1$, and $\alpha_{l-1}^\star = \alpha_l$.
  The special nodes are α_1, α_{l-1}, and α_l.
- $\mathcal{D} = E_6$.
  $\alpha_1^\star = \alpha_6$, $\alpha_2^\star = \alpha_2$, $\alpha_3^\star = \alpha_5$, and $\alpha_4^\star = \alpha_4$.
  The special nodes are α_1 and α_6.
- $\mathcal{D} = E_7$, E_8, F_4, or G_2.
  $\star$ is trivial.
  Only E_7 has a special node, which is α_7.

3.3. ***Multiplicative groups of CM type.*** From now on let $k = \mathbb{Q}$ and $\Gamma = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Let T_1 and T_2 be algebraic groups over $\mathbb{Q}$ of multiplicative type, not necessarily connected. Then there is a natural bijection $\mathrm{Hom}(T_1, T_2) \simeq \mathrm{Hom}_\Gamma(X_2, X_1)$, where Aut_Γ means Γ-equivariant morphisms for the natural Galois structures on $X_i = X^*(T_i)$. In particular, for T over $\mathbb{Q}$ of multiplicative type, there is a natural isomorphism $\mathrm{Aut}(T) \simeq \mathrm{Aut}_\Gamma(X)$, with $X = X^*(T)$. We let $c_T^* : X \to X$ be the map $c_T^*(\chi) = {}^c\chi$. We say T splits over an extension $K \subset \overline{\mathbb{Q}}$ of $\mathbb{Q}$ if $\mathrm{Aut}(\overline{\mathbb{Q}}/K)$ acts trivially on $X^*(T)$.

Lemma 3.3.1. *If T is a group of multiplicative type that splits over a CM field, then $c_T^* \in \mathrm{Aut}_\Gamma(X)$.*

Proof. Suppose that T splits over $K \subset \overline{\mathbb{Q}}$, a CM field. Let $\chi \in X$. Then ${}^\gamma\chi = \chi$ for any $\gamma \in \mathrm{Aut}(\overline{\mathbb{Q}}/K)$, and thus, ${}^{\gamma_1}\chi = {}^{\gamma_2}\chi$ if $\gamma_1, \gamma_2 \in \Gamma$ have the same restriction to K. For any $\gamma \in \Gamma$, γc and $c\gamma$ have the same restriction to K, and so $c_T^*({}^{\gamma}\chi) = {}^{c}({}^{\gamma}\chi) = {}^{c\gamma}\chi = {}^{\gamma c}\chi = {}^{\gamma}(c_T^*(\chi))$. □

Under the assumptions of the last lemma, we let $c_T : T \to T$ denote the unique involution inducing c_T^* on X. If T_1 and T_2 are groups of multiplicative type which are split over a CM field, and $f : T_1 \to T_2$ is a morphism, then $f \circ c_{T_1} = c_{T_2} \circ f$, because both maps induce the same morphism $X_2 \to X_1$.

Suppose now that T is a group of multiplicative type over $\mathbb{R}$. Using the same procedure, there exists a unique involution $c_T : T \to T$ inducing complex conjugation on characters. If T is defined over $\mathbb{Q}$ and split over a CM field, these definitions are compatible with base change from $\mathbb{Q}$ to $\mathbb{R}$.

Example 3.3.2. For $T = \mathbb{S}$ over $\mathbb{R}$, the map $c_{\mathbb{S}}$ is given by $c_{\mathbb{S}}(z \otimes a) = \bar{z} \otimes a$ for an $\mathbb{R}$-algebra A and $z \otimes a \in (\mathbb{C} \otimes_{\mathbb{R}} A)^{\times}$.

Remark 3.3.3. If T is an anisotropic $\mathbb{R}$-torus (that is, if $T(\mathbb{R})$ is compact), then it is easy to see that ${}^{c}\chi = -\chi$ for any $\chi \in X$ and thus $c_T = \mathrm{inv}_T$ is the opposition involution on T.

3.4. *Involutions taking X to $\overline{X}$.* Let (G, X) be a Shimura datum. Recall that we are assuming that Z^0 splits over a CM field, and hence, we have the conjugation involution $c_{Z^0} : Z^0 \to Z^0$.

Remark 3.4.1. Let $x \in X$. From the fact that $\mathrm{int}(x(i)) : G_{\mathbb{R}}^{\mathrm{ad}} \to G_{\mathbb{R}}^{\mathrm{ad}}$ is a Cartan involution, it follows that $G_{\mathbb{R}}^{\mathrm{ad}}$ is an inner form of an anisotropic group H over $\mathbb{R}$ (that is, $H(\mathbb{R})$ is compact). A similar statement holds for $G_{\mathbb{R}}^{\mathrm{der}}$ (the element $x(i)$ may not belong to $G^{\mathrm{der}}(\mathbb{R})$; however, over $\mathbb{C}$, $\mathrm{int}(x(i))$ can be replaced by $\mathrm{int}(x(i)')$ for some $x(i)' \in (T \cap G^{\mathrm{der}})(\mathbb{C})$). The next lemma is well known.

Lemma 3.4.2. *Let G be a reductive group over $\mathbb{R}$, and assume that it is an inner form of a group H over $\mathbb{R}$ which is anisotropic. Assume furthermore that $T \subset G$ is a maximal torus, and the inner automorphism of $G_{\mathbb{C}}$ defining a cocycle for H is given by* $\mathrm{int}(t_0)$ *for some $t_0 \in T(\mathbb{C})$. Then the following hold.*

(i) $c_T = \mathrm{inv}_T$.

(ii) *For a Borel subgroup $B \supset T_{\mathbb{C}}$, the opposition involution acting on $\Psi_0(G, T, B)$ is given by the $*$-action of c.*

(iii) *The subgroup $c(B) \subset G_{\mathbb{C}}$ is the opposite Borel subgroup of B; that is,*

$$c(B) \cap B = T_{\mathbb{C}}.$$

Proof. By hypothesis, we can choose an isomorphism $\phi : G_{\mathbb{C}} \to H_{\mathbb{C}}$ such that $f : G_{\mathbb{C}} \to G_{\mathbb{C}}$ defined by $f(g) = \phi^{-1}\overline{(\phi(\bar{g}))}$ is an inner automorphism of the form $\mathrm{int}(t_0)$, with $t_0 \in T(\mathbb{C})$. Then there exists a maximal torus $T_H \subset H$ such that $T_{H,\mathbb{C}} = \phi(T_{\mathbb{C}})$, and we let $B_H = \phi(B)$. Since $T_H(\mathbb{R})$ is compact, this implies that $T(\mathbb{R})$ is compact, so (i) follows from Remark 3.3.3. Let $r = \Psi_0(\phi) : \Psi_0(H, T_H, B_H) \to \Psi_0(G, T, B)$ be the induced isomorphism. It is $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$-equivariant for the $*$-actions, as follows from the fact that the forms are inner, and it commutes with $\star$, so it is enough to prove part (ii) when G itself is anisotropic, which is well known. For (iii), the fact that f preserves $T_{\mathbb{C}}$ and B again allows us to reduce to the case of G anisotropic, in which case the statement is well known. □

Remark 3.4.3. In the last lemma, if the group is quasisplit and B is a Borel subgroup defined over $\mathbb{R}$, the inner automorphism will not usually belong to $T(\mathbb{C})$; otherwise, we would have $B = T$. There are quasisplit semisimple groups with $B \neq T$ which are inner forms of anisotropic groups, for example $\mathrm{SU}(n, n)$. In this case, the Cartan involution coming from a certain Shimura datum and special pair will preserve the maximal torus and a Borel subgroup containing it, but not a rational Borel subgroup.

Remark 3.4.4. Suppose that (G, X) is a Shimura datum, and let (T, x) be a special pair. Then $G^{\mathrm{der}}_{\mathbb{R}}$ satisfies all the hypotheses of the previous lemma. Here the inner automorphism defining the cocycle is $\mathrm{int}(x(i)')$ as before. Alternatively, we can work with the adjoint group $G^{\mathrm{ad}}_{\mathbb{R}}$ and $x(i)$.

Remark 3.4.5. Suppose that $\theta : G \to G$ is an involution such that there exists a special pair (T, x) with the property that θ preserves T and induces $c_{T_{\mathbb{R}}}$ on $T_{\mathbb{R}}$. Then $\theta_{\mathbb{R}}(x) = c_{T_{\mathbb{R}}} \circ x = x \circ c_{\mathbb{S}} = \bar{x}$, and thus, $\theta(X) = \overline{X}$.

Lemma 3.4.6. *Let G be a reductive group over $\mathbb{R}$, and $T \subset G$ a maximal torus. If $\theta : G \to G$ is an involution such that $\theta(T) \subset T$ and $\theta|_T = c_T$, then $\theta(B) = c(B) \subset G_{\mathbb{C}}$ for any Borel subgroup $B \supset T_{\mathbb{C}}$.*

Proof. Let $R \subset X = X^*(T)$ denote the set of roots of $(G_{\mathbb{C}}, T_{\mathbb{C}})$. Let R^+ denote the set of positive roots with respect to B. Then $\theta(B)$ is the Borel subgroup whose Lie algebra is $\mathrm{Lie}(T_{\mathbb{C}}) \oplus \bigoplus_{\alpha \in R^+} \mathrm{Lie}(G_{\mathbb{C}})_{\alpha\circ\theta}$. Since $\alpha \circ \theta = {}^c\alpha$, it follows that this is the Lie algebra of $c(B)$, and since both $\theta(B)$ and $c(B)$ are connected, this proves the lemma. □

The construction of involutions taking X to $\overline{X}$ that we will perform will be based on involutions θ which will roughly be as in Remark 3.4.5. By the following proposition, we need to look for opposition involutions on semisimple groups.

Proposition 3.4.7. *Let (G, X) be a Shimura datum, and let $\theta : G \to G$ be an involution of G, such that there exists a special pair (T, x) with the property that θ preserves T and induces $c_{T_{\mathbb{R}}}$ on $T_{\mathbb{R}}$. Then $\theta^{\mathrm{der}} : G^{\mathrm{der}} \to G^{\mathrm{der}}$ is an opposition involution, and $\theta_0 = \theta|_{Z^0} : Z^0 \to Z^0$ is equal to c_{Z^0}.*

Proof. Suppose that θ is an involution with (T, x) as in the statement. To see that $\theta_0 = c_{Z^0}$, it is enough to see that $\theta_{\mathbb{R},0} = c_{Z^0_{\mathbb{R}}}$. Since $Z^0_{\mathbb{R}} \subset T_{\mathbb{R}}$ and $\theta_{\mathbb{R}}|_{T_{\mathbb{R}}} = c_{T_{\mathbb{R}}}$, it follows that $\theta_{0,\mathbb{R}} = c_{Z^0_{\mathbb{R}}}$. Let $T' = T \cap G^{\mathrm{der}}$, let $B \subset G_{\mathbb{C}}$ be a Borel subgroup containing $T_{\mathbb{C}}$, and $B' = B \cap G^{\mathrm{der}}_{\mathbb{C}} \supset T'_{\mathbb{C}}$. Let $\Psi'_0 = \Psi_0(G^{\mathrm{der}}, T', B')$, and let $r = \Psi_0(\theta^{\mathrm{der}}) : \Psi'_0 \to \Psi'_0$ be the induced isomorphism. It is given by $r(\chi) = \chi \circ \mathrm{int}(q) \circ \theta^{\mathrm{der}}|_{T'_{\mathbb{C}}}$ for $\chi \in X' = X^*(T')$, where $q \in G^{\mathrm{der}}(\mathbb{C})$ is such that $\mathrm{int}(q)\theta^{\mathrm{der}}(T'_{\mathbb{C}}, B') = (T'_{\mathbb{C}}, B')$. On the other hand, by Lemma 3.4.2(ii), $\star : \Psi'_0 \to \Psi'_0$ is given by $\chi^\star = {}^c\chi \circ \mathrm{int}(a^{-1})$, where $a \in G^{\mathrm{der}}(\mathbb{C})$ is such that $\mathrm{int}(a)c(T'_{\mathbb{C}}, B') = (T'_{\mathbb{C}}, B')$. By Lemma 3.4.6, we can take $a = q$. Finally, the hypothesis that $\theta^{\mathrm{der}}|_{T'} = c_{T'}$ implies that $\chi^\star = \chi \circ \theta \circ \mathrm{int}(q^{-1})$. Thus, to see that $r(\chi) = \chi^\star$, it is enough to see that $\theta^{\mathrm{der}} \circ \mathrm{int}(q^{-1})$ and $\theta^{\mathrm{der}} \circ \mathrm{int}(\varphi(q))$ induce the same automorphism of $T'_{\mathbb{C}}$, and this follows from the fact that both elements $\theta^{-1}(q)$ and q^{-1} conjugate the Borel pair $(T'_{\mathbb{C}}, B')$ to the same Borel pair. $\square$

The following proposition is a partial converse and the main result of this section. Since our construction will be explicit using the classification of semisimple groups, we need to work with either the derived group or the adjoint group. The idea is to construct an involution on G taking X to $\overline{X}$ by extending an opposition involution on G^{der}. Ideally we would want the involution to be as in Remark 3.4.5, but it is enough to consider a weaker hypothesis, as stated in the proposition. Recall the notation from Section 2.1. Suppose that for each i, $S_i \subset H_i$ is a maximal torus, and let $\widetilde{T}_i = \mathrm{Res}_{F_i/\mathbb{Q}} S_i \subset \widetilde{G}_i$, $T_i \subset G_i$ its image in G_i, $T' \subset G^{\mathrm{der}}$ the image of their product, and $T = Z^0 T'$. Note that $T^{\mathrm{ad}}_{\mathbb{R}} = T'^{\mathrm{ad}}_{\mathbb{R}} = \prod_{i,v} S^{\mathrm{ad}}_{i,v}$, where $S_{i,v} \subset H_{i,v}$ and $S^{\mathrm{ad}}_{i,v}$ is its image in $H^{\mathrm{ad}}_{i,v}$.

Proposition 3.4.8. *Suppose that $\theta_i : H_i \to H_i$ is an opposition involution for each i. Suppose moreover that $\theta_i(S_i) = S_i$ and $\theta^{\mathrm{ad}}_{i,v}|_{S^{\mathrm{ad}}_{i,v}} = c_{S^{\mathrm{ad}}_{i,v}}$ for every i and $v \in I_{i,nc}$. Finally, assume that there exists $x \in X$ such that x^{ad} factors through $T^{\mathrm{ad}}_{\mathbb{R}}$. Then there exists an involution $\theta : G \to G$ such that $\theta(X) = \overline{X}$.*

Proof. For each i, the involution $\mathrm{Res}_{F_i/\mathbb{Q}} \theta_i$ defines an opposition involution of $\widetilde{G}_i$. Moreover, the kernel K_i of the projection $\widetilde{G}_i \to G_i$ is contained in the center of $\widetilde{G}_i$. By Lemma 3.1.6, $\mathrm{Res}_{F_i/\mathbb{Q}} \theta_i$ induces $x \mapsto x^{-1}$ on the center. In particular, it preserves K_i and induces an opposition involution on G_i. Similarly, the product of these involutions defines an opposition involution $\theta' : G^{\mathrm{der}} \to G^{\mathrm{der}}$. Let $q : Z^0 \times G^{\mathrm{der}} \to G$ be the natural isogeny. We can look at the product involution $\theta' \times c_{Z^0} : G^{\mathrm{der}} \times Z^0 \to G^{\mathrm{der}} \times Z^0$. We claim that this preserves $\ker(q)$, and thus, it induces an involution on G. To show this, we can work with $\mathbb{C}$-points. The kernel consists of pairs (g, z) such that $zg = 1$, so we need to check that if (g, z) is such a pair, then $\theta'(g)c_{Z^0}(z) = 1$. The element $g = z^{-1}$ belongs to $Z^0 \cap G^{\mathrm{der}} \subset Z_{G^{\mathrm{der}}}$. The maps $c_{Z^0} : Z^0 \to Z^0$ and $c_{Z_{G^{\mathrm{der}}}} : Z_{G^{\mathrm{der}}} \to Z_{G^{\mathrm{der}}}$ are equal on $Z^0 \cap G^{\mathrm{der}}$, and by part (i) of Lemma 3.4.2, $c_{Z_{G^{\mathrm{der}}}} = \mathrm{inv}_{Z_{G^{\mathrm{der}}}}$, so $c_{Z^0}(z) = z^{-1}$. On the other hand, by Lemma 3.1.6, θ' induces $\mathrm{inv}_{Z_{G^{\mathrm{der}}}}$ on $Z_{G^{\mathrm{der}}}$, and so $\theta'(g) = g^{-1} = z$. This proves that there exists a (unique) involution $\theta : G \to G$ such that $\theta^{\mathrm{der}} = \theta'$ and $\theta_0 = c_{Z^0}$.

We also have that θ preserves T and $\theta_{\mathbb{R}}^{\mathrm{ad}} = \prod_{i,v} \theta_{i,v}^{\mathrm{ad}}$. Now, we know that there exists $x \in X$ such that x^{ad} factors through $T_{\mathbb{R}}^{\mathrm{ad}}$. Let $y = \theta_{\mathbb{R}}(x)$. Then $y^{\mathrm{ad}} = \theta_{\mathbb{R}}^{\mathrm{ad}}(x^{\mathrm{ad}})$. For $v \in I_{i,nc}$, we have $\theta_{i,v}^{\mathrm{ad}}(x_{i,v}) = \overline{x_{i,v}}$ because $\theta_{i,v}^{\mathrm{ad}}$ induces $c_{S_{i,v}^{\mathrm{ad}}}$ and $x_{i,v}$ factors through $S_{i,v}^{\mathrm{ad}}$. For $v \in I_{i,c}$, $x_{i,v} = 1$. Thus, $y^{\mathrm{ad}} = \bar{x}^{\mathrm{ad}}$. Also, since $\theta_{\mathbb{R}}$ induces $c_{Z_{\mathbb{R}}^0}$ on $Z_{\mathbb{R}}^0$, and $q : Z^0 \to G^{\mathrm{ab}}$ is an isogeny, it follows that $\theta_{\mathbb{R}}$ induces $c_{G_{\mathbb{R}}^{\mathrm{ab}}}$ on $G_{\mathbb{R}}^{\mathrm{ab}}$. From this it follows that y and $\bar{x}$ have the same projections to $G_{\mathbb{R}}^{\mathrm{ad}}$ and to $G_{\mathbb{R}}^{\mathrm{ab}}$, and thus, $y = \bar{x}$ (see for instance the proof of Proposition 5.7 of [Milne 2005]). Since $y = \theta_{\mathbb{R}}(x)$, this shows that $\theta(X) = \overline{X}$. □

4. Involutions on classical semisimple groups

In this section, we make use of several results regarding the classification of semisimple algebraic groups over totally real fields. For notation and terminology regarding algebras with involutions and their associated groups, we freely follow our main reference [Knus et al. 1998]. We are only interested in the explicit classification of groups of type A and D in order to construct our desired involutions on certain Shimura varieties. Furthermore, not all the groups in the general classification appear in the theory of Shimura varieties, so we are only interested in classifying the groups H_i (in the notation of Section 2.1) of type A_l ($l \geq 2$) or D_l ($l \geq 4$ odd) that can occur. Furthermore, in accordance with the previous section, we are also interested in constructing, whenever possible, opposition involutions on these groups.

The following construction regarding quaternion algebras will be used often in the following. Suppose that D is a quaternion division algebra over a number field K. Let $\lambda \in D^\times$ be a pure quaternion (that is, such that $\sigma(\lambda) = -\lambda$, where $\sigma : D \to D$ is the canonical involution), and choose another pure quaternion $\mu \in D^\times$ such that $\lambda\mu = -\mu\lambda$. Then $\{1, \lambda, \mu, \lambda\mu\}$ is a standard basis of D. If we let $L = K(\lambda)$, then L is a maximal subfield of D (a quadratic extension of K). We have an isomorphism of L-algebras $\phi : D \otimes_K L \to M_2(L)$ defined by

$$\phi(\lambda \otimes 1) = \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix}$$

and

$$\phi(\mu \otimes 1) = \begin{pmatrix} 0 & \mu^2 \\ 1 & 0 \end{pmatrix}.$$

Then the isomorphism ϕ sends $L \otimes_K L$ to the subalgebra of diagonal matrices in $M_2(L)$.

Throughout this section, let F be a totally real field and H be an absolutely almost simple, simply connected algebraic group over F. We let $\mathcal{D}$ be the Dynkin diagram of $H_{\overline{F}}$ (where $\overline{F}$ is some algebraic closure of F). We let $I = \mathrm{Hom}(F, \mathbb{R})$, $I_c = \{v \in I : H_v^{\mathrm{ad}}(\mathbb{R}) \text{ is compact}\}$, and let I_{nc} be its complement in I.

4.1. ***Groups of type A_l ($l \geq 2$).*** Suppose that $\mathcal{D} = A_l$ with $l \geq 2$. Then there exists a quadratic étale extension K/F (so K/F is a quadratic extension of fields, or $K = F \times F$), and a central simple algebra B over K, of degree $l+1$, endowed with an involution $\tau : B \to B$ of the second kind (that is, inducing ι on K, where ι is the nontrivial automorphism of K which fixes F) such that $H = \mathbf{SU}(B, \tau)$ [Knus et al. 1998, Theorem 26.9]. If H is one of the H_i as above, then K must be a field. Indeed, if otherwise, then $H \simeq \mathbf{SL}_1(A)$ for some central simple algebra A over F of degree $l+1$. For each $v \in \mathrm{Hom}(F, \mathbb{R})$, we have

$$A_v = A \otimes_{F,v} \mathbb{R} \simeq M_{l+1}(\mathbb{R}) \quad \text{or} \quad A_v \simeq M_{(l+1)/2}(\mathbb{H}).$$

In both cases, it follows that H_v is an inner form of $\mathrm{SL}_{l+1,\mathbb{R}}$, so the $*$-action of c is trivial (a condition that does not depend on the Borel pair), and thus, it cannot be the opposition involution because $l \geq 2$. From this and Lemma 3.4.2 it follows that H cannot occur as one of the factors H_i. Thus, we have proved that K must be a field. Moreover, a similar argument implies that K must be totally imaginary, that is, K/F is a CM extension. The adjoint group H^{ad} is $\mathbf{PGU}(B, \tau)$.

We can then write $B = \mathrm{End}_D(V)$ for some central division algebra D over K, endowed with an involution $J : D \to D$ of the second kind, whose action we denote by $d \mapsto d^J$, and a finite-dimensional right D-vector space V. There is a nondegenerate hermitian form $h : V \times V \to D$ inducing the involution $\tau : B \to B$. The pair (V, h) is called a hermitian space over D.

Suppose that $\theta : H \to H$ is an opposition involution. There is a natural isomorphism between $\mathrm{Aut}(H)$ and the group of F-algebra automorphisms of B that commute with τ [Knus et al. 1998, Theorem 26.9], and thus, there exists such an automorphism $\gamma : B \to B$ of order 2, inducing θ. If $\gamma|_K$ is the identity map on K, then $\gamma = \mathrm{int}(b_0)$ for some $b_0 \in B^\times$ by the Skolem–Noether theorem, and b_0 is moreover a similitude for τ. The induced map $\theta : H \to H$ would thus be an inner automorphism, inducing the identity map on the Dynkin diagram, but the opposition involution on A_l is nontrivial for $l \geq 2$. Hence, $\gamma|_K$ must be ι. Let $\bar{B}$ and $\bar{D}$ denote the K-algebras B and D with ι-conjugate structure. Thus, $\gamma : B \to \bar{B}$ is a K-algebra isomorphism. We let $\mathrm{Br}(K)$ be the Brauer group of K and $[B] = [D] \in \mathrm{Br}(K)$ be the class of B in it. Then $[D] = [B] = [\bar{B}] = [\bar{D}]$, which implies that there must exist a ring automorphism $\alpha : D \to D$ inducing ι on K.

Proposition 4.1.1. *Let D be a central division algebra over a CM extension K/F of number fields, endowed with an involution $J : D \to D$ of the second kind. Then the following are equivalent*:

(a) *$D = K$ or D is a quaternion division algebra over K.*

(b) *The order of* $[D] \in \mathrm{Br}(K)$ *is* 1 *or* 2.

(c) *There exists a ring automorphism $\alpha : D \to D$ inducing ι on K.*

Moreover, in this case, α is unique up to composition with an inner automorphism of D. Furthermore, it can be chosen to have order 2 and such that $\alpha J = J\alpha$ is either id_D *if $D = K$ or the canonical involution if D is a quaternion division algebra.*

Proof. The fact that (a) implies (b) in the quaternion algebra case follows from the existence of the canonical involution on D, which gives an isomorphism $D \to D^{\mathrm{op}}$, so $[D] = [D]^{-1}$. To see that (b) implies (a), use [Scharlau 1985, 10.2.3].

Now suppose that (a) is true. If $D = K$, then take $\alpha = \iota$. If D is a quaternion division algebra, let $\sigma : D \to D$ be its canonical involution, and take $\alpha = J\sigma = \sigma J$ (they commute because $J\sigma J$ is a symplectic involution of the first kind on D, and hence equal to σ).

Finally, suppose that $\alpha : D \to D$ is as in (c). The involution $J : D \to D$ induces an isomorphism $D \to \bar{D}^{\mathrm{op}}$, where $\bar{D}$ is the conjugate algebra $\lambda \cdot d = \iota(\lambda)d$ for $\lambda \in K$. Similarly, α induces an isomorphism $D \to \bar{D}$, and thus, in the end we have an isomorphism $D \to D^{\mathrm{op}}$, which implies that the order of $[D]$ is 1 or 2.

The uniqueness of α up to inner automorphism follows because if β is another such automorphism, then $\alpha\beta^{-1} : D \to D$ is a K-linear automorphism and hence inner by the Skolem–Noether theorem. □

Remark 4.1.2. Suppose that D is a quaternion division algebra. Under the conditions of the previous proposition, there exists a unique quaternion algebra $D_0 \subset D$ over F such that $D = D_0 \otimes_F K$ and $J = \sigma_0 \otimes_F \iota$, where σ_0 is the canonical involution of D_0 [Knus et al. 1998, §II.22]. Then the map α constructed in the proof is $\alpha = \mathrm{id}_{D_0} \otimes_F \iota$. We define the *canonical conjugation* $\alpha : D \to D$ (attached to J or D_0) to be $\alpha = \mathrm{id}_{D_0} \otimes_F \iota$. If $D = K$, we also call $\alpha = \iota$ the canonical conjugation.

Thus, we have shown that if there exists $\theta : H \to H$ an opposition involution, then either $D = K$ (and $J = \iota$) or D is a quaternion division algebra (and $J = \sigma_0 \otimes_F \iota$). Conversely, suppose that $D = K$ or D is a quaternion division algebra. We will construct suitable opposition involutions under an additional assumption.

Remark 4.1.3. Suppose that $D = K$ (and $D_0 = F$) or D is a quaternion division algebra. Let $I_s \subset \mathrm{Hom}(F, \mathbb{R})$ be the subset of places $v \in I = \mathrm{Hom}(F, \mathbb{R})$ such that $D_{0,v} = D_0 \otimes_{F,v} \mathbb{R}$ is split, and let $I_{ns} \subset I$ be its complement. We let $I_c \subset I$ be the subset of places v such that $H_v^{\mathrm{ad}}(\mathbb{R})$ is compact, and I_{nc} its complement. The group $H_{\mathbb{R}}$ can be written as a product of special unitary groups $\prod_{v \in I} \mathrm{SU}(p_v, q_v)$, and the compact places are exactly the places where $p_v q_v = 0$.

Definition 4.1.4. We say that a hermitian space (V, h) over D (where $D = K$ or a quaternion division algebra) is *strongly hermitian* if there exists an h-orthogonal D-basis $\beta = \{v_1, \dots, v_n\}$ of V such that $h(v_i, v_i) \in K^\times$ for all i. In the quaternion algebra case, we ask furthermore that $I_{ns} \subset I_c$.

Remark 4.1.5. A hermitian space over $D = K$ is always strongly hermitian.

Remark 4.1.6. The existence of the basis β in Definition 4.1.4 is what allows us to explicitly construct an opposition involution $\theta : H \to H$. In the quaternion algebra case, this involution will define an involution $\theta_v : H_v \to H_v$, and we want this to induce complex conjugation on S_v^{ad} when $v \in I_{nc}$. The involution that we construct does not satisfy this at places $v \in I_{ns}$ (see Remark 4.1.9). Since we only care about noncompact places, we make the assumption $I_{ns} \subset I_c$.

Suppose that (V, h) is strongly hermitian, and let β be a basis as in the definition. Let $I : V \to V$ be the α-semilinear isomorphism obtained by applying α to the coordinates of elements of V with respect to the basis β (this map is inspired by the constructions of [Taylor 2012]). Then $h(I(x), I(y)) = \alpha(h(x, y))$. Let $\theta : H \to H$ be given as $\theta_A(g) = I_A g I_A$ for an F-algebra A and a $D \otimes_F A$-linear automorphism g of $V \otimes_F A$. Let $L \subset D$ be a maximal subfield. More precisely, if $D = K$, then $L = K$, and if D is a quaternion division algebra, take $L = K(\lambda)$, where λ is a pure quaternion in D_0. Let $S = S_{L,\beta}$ be the subgroup of H defined as follows. For an F-algebra A, $H(A) \subset \mathrm{Aut}_{D \otimes_F A}(V \otimes_F A)$, and we let

$$S(A) = \{h \in H(A) : h(v_i \otimes 1) = (v_i \otimes 1)\lambda_i \text{ for some } \lambda_i \in (L \otimes_F A)^\times \ (i = 1, \dots, n)\}.$$

This is a maximal torus in H.

Proposition 4.1.7. *With the above hypotheses, the following statements are true:*

(a) *The involution $\theta : H \to H$ is an opposition involution.*

(b) *We have $\theta(S) = S$ and for every $v \in I_{nc}$, $\theta_v : H_v \to H_v$ induces c_{S_v} on S_v.*

In particular, $\theta_v^{\mathrm{ad}} : H_v^{\mathrm{ad}} \to H_v^{\mathrm{ad}}$ induces $c_{S_v^{\mathrm{ad}}}$ on S_v^{ad} for $v \in I_{nc}$.

Proof. For part (a), it suffices to see that $\theta_L : H_L \to H_L$ is an opposition involution. We can identify H_K with $\mathrm{SL}_{V/D}$, where $\mathrm{SL}_{V/D}(A)$ consists, for a K-algebra A, of the $D \otimes_K A$-linear automorphisms of $V \otimes_K A$ with reduced norm 1. Using the basis β, we can further identify $\mathrm{SL}_{V/D}(A) \cong \mathrm{SL}_n(D \otimes_K A)$. Let $Q \in \mathrm{GL}_n(K)$ be the matrix of h with respect to β. Then it is easy to see that $\theta_A : \mathrm{SL}_n(D \otimes_K A) \to \mathrm{SL}_n(D \otimes_K A)$ is explicitly given by the formula

$$\theta_A(X) = Q^{-1}({}^t X^{-1})^\sigma Q,$$

where $\sigma : D \to D$ is the canonical involution of D if D is a quaternion division algebra, and $\sigma = \mathrm{id}$ if $D = K$. Note that Q is a diagonal matrix in $\mathrm{GL}_n(K)$.

If $D = K$, we denote by $\phi : D \otimes_K L \to L$ the unique obvious isomorphism. If D is a quaternion division algebra, we take $\phi : D \otimes_K L \to M_2(L)$ to be an isomorphism of L-algebras taking $L \otimes_K L$ to the subalgebra of diagonal matrices in $M_2(L)$, as constructed above (we use for this the pure quaternion $\lambda \in D_0$ and another pure quaternion $\mu \in D_0$ such that $\lambda\mu = -\mu\lambda$). In particular, σ preserves L. The identification $H_K(A) \cong \mathrm{SL}_n(D \otimes_K A)$ sends $S_K(A)$ to the subgroup of matrices in $\mathrm{SL}_n(D\otimes_K A)$ which are diagonal and have entries in $L\otimes_K A$. Since σ preserves L, it follows that θ_K sends the torus S_K to itself. Moreover, if we now extend scalars to L, the map ϕ provides an isomorphism

$$H_L \cong \mathrm{SL}_{nm,L}, \tag{4.1.8}$$

where $S_{nm,L}$ is the usual group of $nm\times nm$ matrices of determinant 1; furthermore, the torus S_L maps to the torus of diagonal matrices in $\mathrm{SL}_{nm,L}$ (so S is indeed a maximal torus, as claimed). If $m=1$, then $\theta_L(X)=Q^{-1t}X^{-1}Q$ for $X\in\mathrm{SL}_{n,L}$. Suppose that $m=2$. Write $Q=\mathrm{diag}(q_1,\dots,q_n)$, and let $\widetilde{Q}=\mathrm{diag}(q_1,\dots,q_n,q_1,\dots,q_n)\in\mathrm{GL}_{2n}(K)$. Write matrices $X\in\mathrm{SL}_{2n,L}$ as blocks

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

with A, B, C, D of size $n\times n$. Then $\theta_L:\mathrm{SL}_{2n,L}\to\mathrm{SL}_{2n,L}$ is explicitly given as

$$\theta_L(X)=\widetilde{Q}^{-1}\begin{pmatrix} {}^tD & -{}^tB\\ -{}^tC & {}^tA\end{pmatrix}\widetilde{Q}.$$

From this explicit expression of θ as an involution of $\mathrm{SL}_{nm,L}$, it is easy to see that it preserves the maximal torus S_L of diagonal matrices and that it induces the opposition involution on the root datum.

For part (b), fix $v\in I_{nc}$. We need to check that if $\chi\in X=X^*(S_v)=\mathrm{Hom}(S_v\times_{\mathbb{R}}\mathbb{C},\mathbb{G}_{\mathrm{m},\mathbb{C}})$, then $\chi\circ\theta_{v,\mathbb{C}}={}^c\chi$. To compute ${}^c\chi$, we need to compute how complex conjugation acts on $H_v(\mathbb{C})$. Choose once and for all an extension $\tau:L\hookrightarrow\mathbb{C}$ of v to L. Using the embedding τ and the isomorphism (4.1.8), we can identify $H_v\times_{\mathbb{R}}\mathbb{C}=H_L\times_{L,\tau}\mathbb{C}\cong\mathrm{SL}_{nm,\mathbb{C}}$. Moreover, the action of c on $H_v(\mathbb{C})\cong\mathrm{SL}_{nm}(\mathbb{C})$ is explicitly given as follows. Let $Q_v=\mathrm{diag}(v(q_1),\dots,v(q_n))\in\mathrm{GL}_n(\mathbb{R})$ and $\widetilde{Q}_v=\mathrm{diag}(v(q_1),\dots,v(q_n),v(q_1),\dots,v(q_n))\in\mathrm{GL}_{2n}(\mathbb{R})$. Let

$$\gamma=\begin{pmatrix} 0 & I_n\\ -I_n & 0\end{pmatrix}.$$

If $m=1$ and $X\in\mathrm{SL}_n(\mathbb{C})$, then $c(X)=Q_v^{-1}X^{*,-1}Q$. If $m=2$ and $X\in\mathrm{SL}_{2n}(\mathbb{C})$, then $c(X)=Q_v^{-1}\gamma X^{*,-1}\gamma^{-1}Q_v$. The last case easily follows from (4.1.8) and the fact that $D_{0,v}$ is split. We can identify $X^*(S_v)$ in the standard way with $\mathbb{Z}^{nm}/L$, where $L=\{(k,k,\dots,k):k\in\mathbb{Z}\}$. It then follows easily from our calculations of the action of c that if $\chi\in X^*(S_v)$ is identified with the class of the tuple $(a_1,\dots,a_n)$ in the case $m=1$, respectively the class of the tuple $(a_1,\dots,a_n,b_1,\dots,b_n)$ in the case $m=2$, then ${}^c\chi$ is identified with $(-a_1,\dots,-a_n)$ or with $(-b_1,\dots,-b_n,-a_1,\dots,-a_n)$, respectively. This, together with our formulas for θ, show that θ_v induces c_{S_v} on S_v, which is what we wanted to prove. $\square$

Remark 4.1.9. When $D_{0,v}$ is not split, there is also an explicit formula for c that involves a matrix γ as above, but γ turns out to be a diagonal matrix. So in this case θ_v does not induce c_{S_v} on S_v. We only care about noncompact places, hence our assumption $I_{ns} \subset I_c$.

Remark 4.1.10. Keep the assumptions and notation as above. For each $v \in I_{nc}$, we will construct a map $y : \mathbb{S} \to H_v^{\mathrm{ad}}$ satisfying Deligne's axioms [1979, §1.2.1] and factoring through S_v^{ad}. Namely, fix $\tau : L \hookrightarrow \mathbb{C}$ an extension of v to L, and let $w = \tau|_K$ (so $w = \tau$ when $D = K = L$). Let $D_w = D \otimes_{K,w} \mathbb{C}$ and $J_w : D_w \to D_w$ be defined by $J_w(d \otimes z) = J(d) \otimes \bar{z}$. The group $H_v(A)$ can be identified, using the basis β, with the group of matrices $X \in \mathrm{GL}_n(D_w \otimes_{\mathbb{R}} A)$ such that ${}^t X^{J_w} Q X = Q$ and $\mathrm{Nrd}(X) = 1$. If $m = 1$, let $\phi_\tau : D_w \to \mathbb{C}$ be the unique isomorphism. If $m = 2$, consider the $\mathbb{C}$-algebra isomorphism $\phi_\tau : D_w \to M_2(\mathbb{C})$ given by

$$\phi_\tau(\lambda \otimes_{K,w} 1) = \begin{pmatrix} \tau(\lambda) 0 & \\ 0 & -\tau(\lambda) \end{pmatrix}, \qquad \phi_\tau(\mu \otimes_{K,w} 1) = \begin{pmatrix} 0 & v(\mu^2) \\ 1 & 0 \end{pmatrix}.$$

As above, for any $\mathbb{R}$-algebra A, this induces an isomorphism $\mathrm{GL}_n(D_w \otimes_{\mathbb{R}} A) \cong \mathrm{GL}_{mn}(\mathbb{C} \otimes_{\mathbb{R}} A)$ taking the subgroup of diagonal matrices with entries in $L_w \otimes_{\mathbb{R}} A$ (where $L_w = L \otimes_{K,w} \mathbb{C}$) to the subgroup of diagonal matrices in $\mathrm{GL}_{mn}(\mathbb{C} \otimes_{\mathbb{R}} A)$. Moreover, the corresponding involution $X \mapsto {}^t X^{J_w}$ gets identified with $X \mapsto \gamma X^* \gamma^{-1}$, where if $m = 1$, $\gamma = I_n$, and if $m = 2$, γ is the hermitian matrix defined by

$$\gamma = \begin{pmatrix} 0 & i I_n \\ -i I_n & 0 \end{pmatrix}$$

if $v(\lambda^2) > 0$, and

$$\gamma = \begin{pmatrix} -v(\mu^2) I_n & 0 \\ 0 & I_n \end{pmatrix}$$

if $v(\lambda^2) < 0$ (note that in this case, we must have $v(\mu^2) > 0$). In this way, we can write

$$H_v(A) \cong \{X \in \mathrm{GL}_{2n}(\mathbb{C} \otimes_{\mathbb{R}} A) : (\gamma X^* \gamma^{-1}) Q' X = Q',\ \det(X) = 1\},$$

where $Q' = Q_v$ if $m = 1$ and $Q' = \widetilde{Q}_v = \mathrm{diag}(v(q_1), \dots, v(q_n), \dots, v(q_1), \dots, v(q_n))$ if $m = 2$. Thus, we can identify H_v with the special unitary group $\mathrm{SU}(\gamma^{-1} Q')$ of the hermitian matrix $\gamma^{-1} Q'$, and the maximal torus S_v is the torus of diagonal matrices. Note that H_v^{ad} is also the adjoint group of the similitude unitary group $\mathrm{GU}(\gamma^{-1} Q')$. We define $y' : \mathbb{S} \to \mathrm{GU}(\gamma^{-1} Q')$ as follows. For an $\mathbb{R}$-algebra A and $z \in \mathbb{S}(A)$, let

$$y'_A(z) = \begin{pmatrix} \mathrm{diag}(y'_A(z)_1, \dots, y'_A(z)_n) & 0 \\ 0 & \mathrm{diag}(y'_A(z)_1, \dots, y'_A(z)_n) \end{pmatrix},$$

where $y'_A(z)_i = z$ if $v(q_i) > 0$ and $y'_A(z)_i = \bar{z}$ if $v(q_i) < 0$. We let $y = y'^{\mathrm{ad}} : \mathbb{S} \to H_v^{\mathrm{ad}}$.

Using the explicit computation of $\gamma^{-1}Q'$ in each case, the group $\mathrm{GU}(\gamma^{-1}Q')$ is isomorphic to a similitude unitary group $\mathrm{GU}(p,q)$ of a certain signature (p,q) (furthermore, if $m=2$, in our case where $D_{0,v}$ is split, the signature is always (n,n), so the group H_v is in fact quasisplit). It is then standard that y', and hence y, satisfy Deligne's axioms (see for instance the Appendix of [Milne and Shih 1981]).

4.2. ***Groups of type D_l ($l \geq 4$ odd).*** Suppose that $\mathcal{D}=D_l$ with $l\geq 5$ odd. Then $H=\mathbf{Spin}(B,\tau)$, where B is a central simple algebra over F of degree $2l$ and τ is an orthogonal involution [Knus et al. 1998, Theorem 26.15]. The adjoint group is $H^{\mathrm{ad}}=\mathbf{PGO}^+(B,\tau)$. In order to avoid introducing spin groups, we will work in this section with H^{ad}. Since the map $\mathrm{Aut}(H)\to\mathrm{Aut}(H^{\mathrm{ad}})$ is an isomorphism, an opposition involution on H^{ad} will uniquely lift to an opposition involution on H; moreover, suppose that $S\subset H$ is a maximal torus and the involution on H^{ad} preserves S^{ad} and induces $c_{S_v^{\mathrm{ad}}}$ on S_v^{ad} for every $v\in I_{nc}$. Then the lifted involution on H preserves S and also obviously induces $c_{S_v^{\mathrm{ad}}}$ on S_v^{ad} for every $v\in I_{nc}$. This will allow us to concentrate on H^{ad} and avoid spin groups.

Since F is a number field, it can be shown that $B=\mathrm{End}_D(\Lambda)$, where $D=F$ or a quaternion division algebra over F [Scharlau 1985, §8.2.3], and Λ is a right D-vector space of finite dimension n. Let $m=\deg_F D$. Moreover, the involution $\tau: B\to B$ must be attached to a nondegenerate F-bilinear form $q:\Lambda\times\Lambda\to D$. In the case $D=F$ (where $\dim_F\Lambda=2l$), q is a symmetric bilinear form. In the case that D is a quaternion division algebra (where $\dim_D\Lambda=l$), q is a skew-hermitian form with respect to the canonical involution $\sigma: D\to D$. We will only treat the case where D is a quaternion division algebra. Let $I_s\subset I=\mathrm{Hom}(F,\mathbb{R})$ be the set of $v:F\to\mathbb{R}$ such that $D_v=D\otimes_{F,v}\mathbb{R}$ is split, and let I_{ns} be its complement in I. For $v\in I_s$, the skew-hermitian form q_v on Λ_v defines a nondegenerate symmetric bilinear form b_v over a real vector space W_v of dimension $2n$ [Scharlau 1985], and then we have that $I_c\subset I_s$ is the set of split places where b_v is definite. As in the Appendix of [Milne and Shih 1981] (type $D^{\mathbb{H}}$), we will assume that $I_c=I_s$. We call the pair (Λ,q) a skew-hermitian space over D. Note that $n=l$ is odd.

Let $\beta=\{v_1,\dots,v_n\}$ be a D-basis of Λ, which is q-orthogonal. The group $H^{\mathrm{ad}}=\mathbf{PGO}^+(\Lambda,q)$ can also be seen as the adjoint group of $G=\mathbf{SO}(\Lambda,q)$, where

$$G(A)=\{g\in\mathrm{Aut}_{D\otimes_F A}(\Lambda_A):\mathrm{Nrd}(g)=1 \text{ and } q_A(g(x),g(y))=q_A(x,y) \text{ for all } x,y\in\Lambda_A\}$$

for an F-algebra A. Here $\Lambda_A=\Lambda\otimes_F A$ and Nrd is the reduced norm in $\mathrm{End}_D(\Lambda)$. We let $S'=S'_\beta\subset G$ be the subgroup of G defined as follows. For every $i=1,\dots,n$, let $q_i=q(v_i,v_i)$. This is a pure quaternion in D, and so $L_i=F(q_i)$ is a quadratic field extension of F. For an F-algebra A, let

$$S'(A)=\{g\in G(A): g(v_i\otimes 1)=(v_i\otimes 1)\lambda_i \text{ for some } \lambda_i\in(L_i\otimes_F A)^\times\ (i=1,\dots,n)\}.$$

Then $S' \subset G$ is a maximal torus of G, and it defines maximal tori $S \subset H$ and $S^{\mathrm{ad}} = S'^{\mathrm{ad}} \subset H^{\mathrm{ad}}$.

We will construct involutions on H modeled after our constructions for the case of type A_l. For this we need to make an analogous extra assumption.

Definition 4.2.1. We say that the skew-hermitian space (Λ, q) over D is *strongly skew-hermitian* if there exists a q-orthogonal D-basis $\beta = \{v_1, \dots, v_n\}$ of Λ and an F-automorphism $\alpha : D \to D$ such that $q(v_i, v_j) = -\alpha(q(v_j, v_i))$ and $\alpha^2 = 1$.

Remark 4.2.2. Any automorphism $\alpha : D \to D$ as above must be necessarily inner, of the form $\alpha(d) = rdr^{-1}$ for some $r \in D^\times$ such that $r^2 \in F^\times$. This implies that $r\sigma(r)^{-1} \in F^\times$ as well (because F is the set of elements of D fixed by σ). Moreover, since $q(v_i, v_i) \in D^\times$, r must be a pure quaternion in D. As in the previous case, the existence of the basis β will allow us to construct an explicit involution. The map α plays the role of the canonical conjugation of case A_l.

Suppose that (Λ, q) is strongly hermitian, and let β and $\alpha = \mathrm{int}(r)$ be as in the definition. We then have $\alpha\sigma = \sigma\alpha$. Let $I : \Lambda \to \Lambda$ be the α-semilinear automorphism obtained by applying α to the coefficients of elements of Λ with respect to the basis β. Then $q(I(x), I(y)) = -\alpha(q(x, y))$. Let $\theta : G \to G$ be defined by $\theta_A(g) = I_A g I_A$ for an F-algebra A and a $(D \otimes_F A)$-linear automorphism g of $\Lambda \otimes_F A$.

Let $L = F(r)$, where $r \in D$ is as above. This is again a quadratic extension of F (and a maximal subfield of D). Let S' and S^{ad} be the maximal tori of G and H^{ad} defined above using the basis β.

Proposition 4.2.3. *With the above hypotheses, the following statements are true*:

(a) *The map $\theta : G \to G$ is an opposition involution* (*and hence so is θ^{ad}*).

(b) *We have $\theta(S') = S'$ and for every $v \in I_{nc}$, $\theta_v : G_v \to G_v$ induces $c_{S'_v}$ on S'_v.*

In particular, $\theta_v^{\mathrm{ad}} : H_v^{\mathrm{ad}} \to H_v^{\mathrm{ad}}$ induces $c_{S_v^{\mathrm{ad}}}$ on S_v^{ad} for $v \in I_{nc}$.

Proof. For part (a), it suffices to see that $\theta_E : G_E \to G_E$ is an opposition involution, for a convenient extension E/F. Using the basis β and the isomorphism $\phi : D \otimes_F L \to M_2(L)$ as constructed above, we can identify G_L as follows. Implicit in the construction of ϕ is the choice of a pure quaternion $s \in D$ with $rs = -rs$, and we let $t = v(s^2) \in \mathbb{R}$. Let $q_i = q(v_i, v_i)$. Since $\sigma(q_i) = -q_i$ and $rq_ir^{-1} = -q_i$, we have

$$\phi(r) = \begin{pmatrix} r & 0 \\ 0 & -r \end{pmatrix} \in \mathrm{GL}_2(L)$$

and

$$\phi(q_i) = \begin{pmatrix} 0 & b_i \\ c_i & 0 \end{pmatrix}$$

for some $b_i, c_i \in L$. The image in $M_2(L)$ under ϕ of $L_i \otimes_F L \subset D \otimes_F L$ consists of the matrices in $M_2(L)$ of the form

$$\begin{pmatrix} x & yb_i \\ yc_i & x \end{pmatrix}$$

for some $x, y \in L$. Thus, the induced isomorphism $\phi : M_n(D \otimes_F L) \to M_{2n}(L)$ sends the subalgebra of diagonal matrices $L_1 \otimes_F L \times \cdots \times L_n \otimes_F L$ to the set of matrices in $M_{2n}(L)$ of the form

$$X = \begin{pmatrix} \operatorname{diag}(x_1, \ldots, x_n) & \operatorname{diag}(y_1 b_1, \ldots, y_n b_n) \\ \operatorname{diag}(y_1 c_1, \ldots, y_n c_n) & \operatorname{diag}(x_1, \ldots, x_n) \end{pmatrix} \tag{4.2.4}$$

with $x_i, y_i \in L$. Let

$$\widetilde{Q} = \begin{pmatrix} 0 & \operatorname{diag}(b_1, \ldots, b_n) \\ \operatorname{diag}(c_1, \ldots, c_n) & 0 \end{pmatrix} \in \mathrm{GL}_{2n}(L).$$

Then, for any L-algebra R, writing a matrix $X \in \mathrm{GL}_{2n}(R)$ as $X = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, there is an isomorphism

$$G(R) \cong \left\{ X \in \mathrm{GL}_{2n}(R) \, : \, \begin{pmatrix} {}^tD & -{}^tB \\ -{}^tC & {}^tA \end{pmatrix} \widetilde{Q} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \widetilde{Q}, \ \det(X) = 1 \right\} \tag{4.2.5}$$

that takes the subgroup S' to the subgroup of matrices of the form (4.2.4) in the right-hand side. Note that the equation is equivalent to ${}^tX \widetilde{Q}' X = \widetilde{Q}'$, where

$$\widetilde{Q}' = \begin{pmatrix} \operatorname{diag}(c_1, \ldots, c_n) & 0 \\ 0 & -\operatorname{diag}(b_1, \ldots, b_n) \end{pmatrix}$$

(the matrix $\widetilde{Q}'$ is the matrix of the associated bilinear form [Scharlau 1985, §10.3]). Moreover, if

$$\gamma = \begin{pmatrix} rI_n & 0 \\ 0 & -rI_n \end{pmatrix},$$

then $\theta_R(X) = \gamma X \gamma^{-1}$ for $X \in G(R)$; in block matrix terms,

$$\theta_R \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & -B \\ -C & D \end{pmatrix}.$$

It is clear then that θ preserves S'.

Let E/L be a field extension such that there exist elements $e_i, f_i \in E$ with $e_i^2 = c_i$ and $f_i^2 = b_i$ (for example, take $E = \mathbb{C}$ with a fixed embedding of L). For elements $a_1, \ldots, a_n$, let $\operatorname{adiag}(a_1, \ldots, a_n)$ be the antidiagonal matrix whose $(i, n+1-i)$-th entry is a_i, and let $J_n = \operatorname{adiag}(1, \ldots, 1)$. Let

$$\delta = \begin{pmatrix} \operatorname{adiag}(e_n, \ldots, e_1) & \operatorname{adiag}(-f_n, \ldots, -f_1) \\ \operatorname{diag}(e_1/2, \ldots, e_n/2) & \operatorname{diag}(f_1/2, \ldots, f_n/2) \end{pmatrix} \in \mathrm{GL}_{2n}(E).$$

Then the map $X \mapsto \delta X \delta^{-1}$ sends G_E (viewed inside $\mathrm{GL}_{2n,E}$ via (4.2.5)) to the special orthogonal group SO_{2n} of the matrix J_{2n} over E. The maximal torus S'_E maps to the subgroup of diagonal matrices in SO_{2n}, and θ becomes conjugation by the matrix

$$\delta\gamma\delta^{-1} = \begin{pmatrix} 0 & 2r J_n \\ (r/2) J_n & 0 \end{pmatrix}$$

inside GL_{2n}. We identify in the usual way $X^*(S') \cong \mathbb{Z}^n$. As a Borel subgroup of G_E we take the subgroup B of upper-triangular matrices belonging to G_E. The map θ sends B to the subgroup B^- of lower-triangular matrices. Let J'_{2n} be the matrix obtained from J_{2n} by swapping the rows n and $n+1$. Then it is easy to see that $J'_{2n} \in G(E)$ and sends B^- to B. It follows that $\Psi_0(\theta)(\chi) = \chi \circ \mathrm{int}(J'_{2n}) \circ \theta$ for $\chi \in X^*(S')$. If χ is parametrized by $(a_1, \dots, a_n)$, then $\Psi_0(\theta)(\chi)$ is parametrized by $(a_1, \dots, a_{n-1}, -a_n) = (a_1, \dots, a_n)^\star$ [Bourbaki 2002, Plate IV]. Thus, $\theta : G \to G$ is an opposition involution.

Let $v : F \hookrightarrow \mathbb{R}$, and let $\tau : L \hookrightarrow \mathbb{C}$ be an extension of v to L. If $\tau = \bar{\tau}$, then $\tau(r) \in \mathbb{R}$. Thus, $\tau(r)^2 \in \mathbb{R}_{>0}$, and this implies that D_v is split, so $v \in I_s = I_c$. In part (b), we only care for $v \in I_{nc}$, so suppose from now on that $\tau \neq \bar{\tau}$, so that $\tau(r) \in i\mathbb{R}_{>0}$. By the same reasoning we have that $t = v(s^2) < 0$. We use τ to identify $G_{\mathbb{C}} \cong \mathrm{SO}_{2n}$ as above. We first work out the induced complex conjugation on $G(\mathbb{C}) \cong \mathrm{SO}_{2n}(\mathbb{C})$. Using the isomorphisms $D \otimes_{F,v} \mathbb{C} \simeq (D \otimes_F L) \otimes_{L,\tau} \mathbb{C} \cong M_2(\mathbb{C})$ (the last one coming from ϕ), it is easy to see that complex conjugation on $D \otimes_{F,v} \mathbb{C}$ corresponds to taking a matrix $X \in M_2(\mathbb{C})$ to

$$\begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \overline{X_{22}} & \overline{X_{21}} \\ \overline{X_{12}} & \overline{X_{11}} \end{pmatrix} \begin{pmatrix} t^{-1} & 0 \\ 0 & 1 \end{pmatrix},$$

where $t = v(s^2)$ as above. Note that $q_i \in D \subset D \otimes_{F,v} \mathbb{C}$, so this implies that $t\overline{\tau(c_i)} = \tau(b_i)$ and thus

$$te_i / \bar{f}_i = -f_i / \bar{e}_i. \tag{4.2.6}$$

It follows that the induced complex conjugation on $G(\mathbb{C})$, viewed inside $\mathrm{GL}_{2n}(\mathbb{C})$ as in (4.2.5), is given by

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto c'(X) = \begin{pmatrix} \bar{D} & t\bar{C} \\ t^{-1}\bar{B} & \bar{A} \end{pmatrix}.$$

Finally, we apply conjugation by δ to identify $G_{\mathbb{C}}$ with SO_{2n}. We only need to consider the action of c on diagonal matrices. Let $X = \mathrm{diag}(x_1, \dots, x_n, x_n^{-1}, \dots, x_1^{-1}) \in \mathrm{SO}_{2n}(\mathbb{C})$. Then $c(X) = \delta c'(\delta)^{-1} c'(X) c'(\delta)\delta^{-1}$, and a long but easy direct calculation using (4.2.6) shows that

$$\delta c'(\delta)^{-1} = \begin{pmatrix} 2\,\mathrm{adiag}(e_n/\bar{f}_n, \dots, e_1/\bar{f}_1) & 0 \\ 0 & \frac{1}{2}\,\mathrm{adiag}(f_1/\bar{e}_1, \dots, f_n/\bar{e}_n) \end{pmatrix},$$

and thus,

$$c(\mathrm{diag}(x_1,\dots,x_n,x_n^{-1},\dots,x_1^{-1}))=\mathrm{diag}((\overline{x_1})^{-1},\dots,(\overline{x_n})^{-1},\overline{x_n},\dots,\overline{x_1}).$$

This implies that, if $\chi\in X^*(S')$ is parametrized by $(a_1,\dots,a_n)\in\mathbb{Z}^n$, then ${}^c\chi$ is parametrized by $(-a_1,\dots,-a_n)$. This is also easily seen to be the parameter of $\chi\circ\theta$, which shows that θ_v induces $c_{S'_v}$ on S'_v. □

Remark 4.2.7. Keep the assumptions and notation as above. For each $v\in I_{nc}$, we will construct a map $y:\mathbb{S}\to H_v^{\mathrm{ad}}$ satisfying Deligne's axioms [1979, §1.2.1] and factoring through S_v^{ad}. Recall that $t=v(s^2)$, and let $u=v(r^2)$. Since $v\in I_{nc}$, by our assumptions D_v is not split. This implies that $u<0$ and $t<0$. Let $\psi:D_v\to\mathbb{H}$ be the isomorphism of $\mathbb{R}$-algebras sending $r\otimes 1$ to $\sqrt{-u}e_2$ and $s\otimes 1$ to $\sqrt{-t}e_3$. Here e_1, e_2, e_3, and e_4 are the following elements of $\mathbb{H}$:

$$e_1=I_2,\qquad e_2=\begin{pmatrix} i & 0\\ 0 & -i\end{pmatrix},\qquad e_3=\begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix},\qquad e_4=e_2e_3.$$

As above, we can write $\psi(q_i)=\begin{pmatrix} 0 & y_i\\ -\overline{y_i} & 0\end{pmatrix}$ with $y_i\in\mathbb{C}^\times$. Let

$$T=\begin{pmatrix} 0 & \mathrm{diag}(y_1,\dots,y_n)\\ -\mathrm{diag}(\overline{y_1},\dots,\overline{y_n}) & 0\end{pmatrix}.$$

We then have, for an $\mathbb{R}$-algebra R,

$$G_v(R)\cong\left\{X=\begin{pmatrix} A & B\\ -\bar{B} & \bar{A}\end{pmatrix}\in\mathrm{GL}_{2n}(\mathbb{C}\otimes_{\mathbb{R}}R)\,:\,X^*TX=T,\ \det(X)=1\right\}.\tag{4.2.8}$$

The maximal torus S' corresponds to the subgroup of matrices on the right-hand side where $A=\mathrm{diag}(a_1,\dots,a_n)$ and $B=\mathrm{diag}(b_1y_1,\dots,b_ny_n)$ with $a_i,b_i\in R$. We can actually see H_v as the adjoint group of G'_v, where $G'_v(R)$ is isomorphic to

$$\left\{X=\begin{pmatrix} A & B\\ -\bar{B} & \bar{A}\end{pmatrix}\in\mathrm{GL}_{2n}(\mathbb{C}\otimes_{\mathbb{R}}R)\,:\,X^*TX=\nu(X)T,\ \det(X)=\nu(X)^n\right\}.$$

We define $y':\mathbb{S}\to G'_v$ by the formula

$$y'_R(z)=\begin{pmatrix} \mathrm{Re}(z)I_n & \mathrm{diag}\Big(\frac{\mathrm{Im}(z)}{|y_1|}y_1,\dots,\frac{\mathrm{Im}(z)}{|y_n|}y_n\Big)\\ \mathrm{diag}\Big(-\frac{\mathrm{Im}(z)}{|y_1|}\overline{y_i},\dots,-\frac{\mathrm{Im}(z)}{|y_n|}\overline{y_n}\Big) & \mathrm{Re}(z)I_n\end{pmatrix}$$

for $z\in\mathbb{S}(R)$. Conjugating by a suitable matrix $U\in\mathrm{GL}_{2n}(\mathbb{C})$, we can write $G'_v\cong\mathrm{GO}^*(2n)$ and y becomes the map in the Appendix of [Milne and Shih 1981], so it satisfies Deligne's axioms, and hence also does $y=y'^{\mathrm{ad}}$.

5. Involutions on certain Shimura varieties

In this section we combine all our previous results to prove the existence of descent data on certain Shimura varieties $\mathrm{Sh}(G, X)$. As we said before, we only consider the case where the simple groups H_i are of type A or $D^{\mathbb{H}}$. In the previous section, we constructed opposition involutions on some of these groups, preserving a certain maximal torus S_i and inducing complex conjugation on its characters. Furthermore, we constructed maps $y_{i,v} : \mathbb{S} \to H_{i,v}^{\mathrm{ad}}$ for every $v \in I_{i,nc}$ satisfying Deligne's axioms [1979, §1.2.1], factoring through $S_{i,v}^{\mathrm{ad}}$. We now show that we can always find an element $x \in X$ such that $x_{i,v}$ factors through $S_{i,v}^{\mathrm{ad}}$ for every i and $v \in I_{i,nc}$. The existence of descent data will follow by combining this with Proposition 3.4.8.

5.1. ***Existence of particular elements $x \in X$.*** Let H be an almost simple, simply connected group over $\mathbb{R}$ (to play the role of one of the noncompact $H_{i,v}$). Suppose that there exist morphisms $y : \mathbb{S} \to H^{\mathrm{ad}}$ satisfying Deligne's axioms [1979, §1.2.1]; in particular, H is absolutely almost simple. Let D be the Dynkin diagram of $H_{\mathbb{C}}$ associated with a choice of maximal torus and Borel subgroup. To each $H^{\mathrm{ad}}(\mathbb{R})$-conjugacy class Y of morphisms y as above, we can attach a special node $s_Y \in D$, and $s_Y = s_{Y'}$ if and only if $Y = Y'$.

Lemma 5.1.1. *Under the above conditions, there exist at most two $H(\mathbb{R})$-conjugacy classes Y of morphisms satisfying Deligne's axioms* [1979, §1.2.1]. *Moreover, given such a conjugacy class Y, any morphism satisfying these axioms must belong to either Y or Y^{-1}.*

Proof. Suppose first that D is not of type A_l. This case is easy because there are not too many special nodes. Indeed, assume first that $H(\mathbb{R})$ is connected, and fix Y one of the conjugacy classes. Then $s_{Y^{-1}} = s_Y^\star \neq s_Y$ [Deligne 1979, §1.2.8], and hence, Y^{-1} and Y are two distinct conjugacy classes. Suppose that Z is a third conjugacy class, that is, s_Z is neither equal to s_Y nor to $s_Y^\star$. Again by [Deligne 1979, §1.2.8], $s_Z \neq s_Z^\star$, and thus, we have four distinct special nodes s_Y, $s_Y^\star$, s_Z, and $s_Z^\star$. There is no connected Dynkin diagram with four special nodes which is not of type A_l, and thus, this is a contradiction. If $H(\mathbb{R})$ is not connected, then $s_Y = s_Y^\star$. If Z is another conjugacy class, then again by [op. cit.] we must have $s_Z = s_Z^\star$. But for any connected Dynkin diagram, there is at most one special node which is fixed under the opposition involution, and thus $Z = Y$.

Suppose now that H is of type A_l with $l \geq 2$, so $H = \mathrm{SU}(p, q)$ for some nonzero pair of integers p, q such that $p+q = l+1$. The isomorphism $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \simeq \mathbb{C} \times \mathbb{C}$ given by $z \otimes a \mapsto (za, \bar{z}a)$ induces by projection on the first coordinate an isomorphism $H_{\mathbb{C}} \simeq \mathrm{SL}_{l+1,\mathbb{C}}$; fix the usual Borel pair here to define the Dynkin diagram. Define a morphism

$$y_0 : \mathbb{S} \to H^{\mathrm{ad}} = \mathrm{PGU}(p, q)$$

with $y_0(z)$ being the class of the matrix

$$\begin{pmatrix} zI_p & 0 \\ 0 & \bar{z}I_q \end{pmatrix}.$$

Then y_0 satisfies Deligne's axioms [1979, §1.2.1], and the special node s_0 attached to its $H^{\mathrm{ad}}(\mathbb{R})$-conjugacy class Y_0 is α_p. From the conjugate map $\bar{y}_0 = y_0^{-1}$ we get the special node α_q associated with Y_0^{-1}. If Y is another conjugacy class, say with special node α_t, then there would be an isomorphism $\mathrm{PGU}(p, q) \cong \mathrm{PGU}(t, l+1-t)$ sending Y_0 or Y_0^{-1} to Y. In particular, $t = p$ or $t = q$, and we conclude that there are at most two possible conjugacy classes of morphisms satisfying Deligne's axioms for the fixed form $\mathrm{PGU}(p, q)$ of $\mathrm{PGL}_{l+1,\mathbb{C}}$ (and there are exactly two in all cases except when $p = q$, when there is only one). □

Going back to our general Shimura datum (G, X), for each i, let $S_i \subset H_i$ be a maximal torus, $\widetilde{T}_i = \mathrm{Res}_{F_i/\mathbb{Q}} S_i \subset \widetilde{G}_i$, $T_i \subset G_i$ its image in G_i, $T' \subset G^{\mathrm{der}}$ the image of their product, and $T = Z^0 T'$. Note that $T_{\mathbb{R}}^{\mathrm{ad}} = T_{\mathbb{R}}^{\prime\mathrm{ad}} = \prod_{i,v} S_{i,v}^{\mathrm{ad}}$, where $S_{i,v} \subset H_{i,v}$ and $S_{i,v}^{\mathrm{ad}}$ is its image in $H_{i,v}^{\mathrm{ad}}$.

Lemma 5.1.2. *Suppose that $T \subset G$ is the maximal torus defined above. Suppose that for each $v \in I_{i,nc}$, there exists a morphism $y_{i,v} : \mathbb{S} \to H_{i,v}^{\mathrm{ad}}$ satisfying axioms* [Deligne 1979, §1.2.1] *and factoring through $S_{i,v}^{\mathrm{ad}}$. Then there exists an element $x \in X$ such that x^{ad} factors through $T_{\mathbb{R}}^{\mathrm{ad}}$.*

Proof. Let $z \in X$ be an arbitrary element. The previous lemma implies that $z_{i,v}$ is $H_{i,v}^{\mathrm{ad}}(\mathbb{R})$-conjugate to a map $y_{i,v} : \mathbb{S} \to S_{i,v}^{\mathrm{ad}}$. Thus, we can write $z_{i,v} = u_{i,v} \cdot y_{i,v}$ for $u_{i,v} \in H_{i,v}^{\mathrm{ad}}(\mathbb{R})$. We claim that, after possibly changing the $y_{i,v}$, we can arrange for $u_{i,v}$ to be in $H_{i,v}^{\mathrm{ad}}(\mathbb{R})^+$. Indeed, if $u_{i,v}$ is not in that connected component, then in particular $H_{i,v}^{\mathrm{ad}}(\mathbb{R})$ is not connected, and thus, there is only one conjugacy class in question, with two connected components, one containing $z_{i,v}$ and the other one containing $y_{i,v}$. Thus, we only need to replace $y_{i,v}$ with $y_{i,v}^{-1}$, which also factors through $S_{i,v}^{\mathrm{ad}}$. For $v \in I_{i,c}$, let $u_{i,v} = 1$. It follows that $u = (u_{i,v}) \in G^{\mathrm{ad}}(\mathbb{R})^+$, and thus, there exists $g \in G(\mathbb{R})$ lifting u. Let $x = g^{-1} \cdot z \in X$, so that $x^{\mathrm{ad}} = (y_{i,v})$, which factors through $T_{\mathbb{R}}^{\mathrm{ad}}$ as desired. □

5.2. *The main theorem.* In this subsection we put all the ingredients together to obtain the main theorem on the existence of involutions of G taking X to $\overline{X}$.

Definition 5.2.1. The Shimura datum (G, X) is said to be *strongly of type* $(AD^{\mathbb{H}})$ if each of the groups H_i is either of type A_l with $l \geq 2$ and attached to a strongly hermitian space (as in Definition 4.1.4), or of type D_l with $l \geq 5$ odd and attached to a strongly skew-hermitian space (as in Definition 4.2.1).

For example, a Shimura variety defined by a similitude unitary group attached to a hermitian space over a CM field is strongly of type $(AD^{\mathbb{H}})$. Note however that the definition only restricts the semisimple part of G.

Theorem 5.2.2. *Suppose that* (G, X) *is strongly of type* $(AD^{\mathbb{H}})$. *Then there exists an involution* $\theta : G \to G$ *such that* $\theta(X) = \overline{X}$, *and hence, there exists a model of* $\mathrm{Sh}(G, X)$ *over* E^+ *as in Theorem 2.3.1.*

Proof. In Subsections 4.1 and 4.2, we constructed for every i, an opposition involution $\theta_i : H_i \to H_i$ and a maximal torus $S_i \subset H_i$ such that $\theta_i(S_i) = S_i$ and $\theta_{i,v}^{\mathrm{ad}}$ induces $c_{S_{i,v}^{\mathrm{ad}}}$ for every $v \in I_{i,nc}$. Moreover, by Remarks 4.1.10 and 4.2.7, for every i and $v \in I_{i,nc}$, there is a map $y_{i,v} : \mathbb{S} \to H_{i,v}^{\mathrm{ad}}$ satisfying Deligne's axioms [1979, §1.2.1], and factoring through $S_{i,v}^{\mathrm{ad}}$. The result then follows by combining Proposition 3.4.8 and Lemma 5.1.2. □

Remark 5.2.3. The conclusion of the previous theorem holds in other cases as well. For instance, if G is adjoint and there exists an opposition involution $\theta : G \to G$ (which is always the case if G is also quasisplit, for example), then by the adjointness of G, we conclude that $\theta(X) = \overline{X}$. On the other hand, the cases that we considered in this paper are concretely given by simple algebras, and thus are intimately related to moduli problems, even though we do not use the moduli interpretation explicitly. It is an interesting problem to consider factors of other types, for instance of type E_6, and analyze whether it is possible to construct opposition involutions with the desired properties in some of these cases. We plan to investigate this problem in the future.

Acknowledgements

The first author thanks the Erwin Schrödinger International Institute for Mathematics and Physics for its support during a visit when a part of the work for this project was undertaken. A substantial part of this work was carried out while the second author was a guest at the Max Planck Institute for Mathematics in Bonn, Germany. It is a pleasure to thank the Institute for its hospitality and the excellent working conditions. The authors would also like to thank the anonymous referees for many useful comments and suggestions which helped us improve the paper.

References

[Borovoĭ 1983] M. V. Borovoĭ, "Langlands' conjecture concerning conjugation of connected Shimura varieties", *Selecta Math. Soviet.* **3**:1 (1983), 3–39. MR

[Borovoĭ 1987] M. V. Borovoĭ, "The group of points of a semisimple group over a totally real-closed field", pp. 142–149 in *Problems in group theory and homological algebra*, edited by A. L. Onishchik, Yaroslavl State University, 1987. In Russian. MR

[Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Ergebnisse der Mathe. und ihrer Grenzgebiete (3) **21**, Springer, 1990. MR Zbl

[Bourbaki 2002] N. Bourbaki, *Lie groups and Lie algebras, Chapters 4–6*, Elements of Math. **9**, Springer, 2002. MR Zbl

[Deligne 1979] P. Deligne, "Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques", pp. 247–289 in *Automorphic forms, representations and L-functions* (Corvallis, OR, 1977), vol. 2, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, American Math. Society, 1979. MR Zbl

[Deligne 1982] P. Deligne, "Motifs et groupes de Taniyama", pp. 261–279 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. **900**, Springer, 1982. MR Zbl

[Demazure 1965/66] M. Demazure, "Automorphismes des groupes réductifs", exposé XXIV in *Schémas en groupes, Fasc. 7: Exposés 23 à 26* (Séminaire de Géométrie Algébrique de l'Institut des Hautes Études Scientifiques, 1963/64), edited by M. Demazure and A. Grothendieck, IHES, Paris, 1965/66. Reprinted as pp. 323–409 in *Schémas en groupes, Tome III: Structure des schémas en groupes réductifs*, Lecture Notes in Math. **153**, Springer, 1970. MR Zbl

[Knus et al. 1998] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, AMS Colloquium Publications **44**, American Math. Society, 1998. MR Zbl

[Langlands 1979] R. P. Langlands, "Automorphic representations, Shimura varieties, and motives: ein Märchen", pp. 205–246 in *Automorphic forms, representations and L-functions* (Corvallis, OR, 1977), vol. 2, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, American Math. Society, 1979. MR Zbl

[Milne 1983] J. S. Milne, "The action of an automorphism of $\mathbb{C}$ on a Shimura variety and its special points", pp. 239–265 in *Arithmetic and geometry*, vol. I, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, 1983. MR Zbl

[Milne 1990] J. S. Milne, "Canonical models of (mixed) Shimura varieties and automorphic vector bundles", pp. 283–414 in *Automorphic forms, Shimura varieties, and L-functions, I* (Ann Arbor, MI, 1988), edited by L. Clozel and J. S. Milne, Perspect. Math. **10**, Academic, 1990. MR Zbl

[Milne 1999] J. S. Milne, "Descent for Shimura varieties", *Michigan Math. J.* **46**:1 (1999), 203–208. MR Zbl

[Milne 2005] J. S. Milne, "Introduction to Shimura varieties", pp. 265–378 in *Harmonic analysis, the trace formula, and Shimura varieties* (Toronto, 2003), edited by J. Arthur et al., Clay Math. Proc. **4**, American Math. Society, 2005. MR Zbl

[Milne and Shih 1981] J. S. Milne and K.-y. Shih, "The action of complex conjugation on a Shimura variety", *Ann. of Math.* (2) **113**:3 (1981), 569–599. MR Zbl

[Milne and Shih 1982a] J. S. Milne and K.-y. Shih, "Conjugates of Shimura varieties", pp. 280–356 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. **900**, Springer, 1982. MR Zbl

[Milne and Shih 1982b] J. S. Milne and K.-y. Shih, "Langlands's construction of the Taniyama group", pp. 229–260 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. **900**, Springer, 1982. MR Zbl

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der mathe. Wissenschaften **270**, Springer, 1985. MR Zbl

[Springer 1979] T. A. Springer, "Reductive groups", pp. 3–27 in *Automorphic forms, representations and L-functions* (Corvallis, OR, 1977), vol. 1, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, American Math. Society, 1979. MR Zbl

[Taylor 2012] R. Taylor, "The image of complex conjugation in l-adic representations associated to automorphic forms", *Algebra Number Theory* **6**:3 (2012), 405–435. MR Zbl

[Weil 1956] A. Weil, "The field of definition of a variety", *Amer. J. Math.* **78** (1956), 509–524. MR Zbl

Communicated by Shou-Wu Zhang
Received 2016-02-21 Revised 2017-03-30 Accepted 2017-04-29

blasius@math.ucla.edu *Mathematics Department, University of California, Los Angeles, Los Angeles, CA, United States*

lguerb@gmail.com *Department of Mathematics, University College London, London, United Kingdom*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2323

A subspace theorem for subvarieties

Min Ru and Julie Tzu-Yueh Wang

We establish a height inequality, in terms of an (ample) line bundle, for a sum of subschemes located in ℓ-subgeneral position in an algebraic variety, which extends a result of McKinnon and Roth (2015). The inequality obtained in this paper connects the result of McKinnon and Roth (the case when the subschemes are points) and the results of Corvaja and Zannier (2004), Evertse and Ferretti (2008), Ru (2017), and Ru and Vojta (2016) (the case when the subschemes are divisors). Furthermore, our approach gives an alternative short and simpler proof of McKinnon and Roth's result.

1. Introduction and statements

McKinnon and M. Roth [2015] introduced the approximation constant $\alpha_x(L)$ to an algebraic point x on an algebraic variety V with an ample line bundle L. The invariant $\alpha_x(L)$ measures how well x can be approximated by rational points on V with respect to the height function associated to L. They showed that $\alpha_x(L)$ is closely related to the Seshadri constant $\epsilon_x(L)$ measuring the local positivity of L at x. They also showed that the invariant $\alpha_x(L)$ can be computed through another invariant $\beta_x(L)$ in the height inequality they established (see Theorem 5.1 and Theorem 6.1 in [McKinnon and Roth 2015]). By computing the Seshadri constant $\epsilon_x(L)$ for the case of $V = \mathbb{P}^1$, their result recovers Roth's theorem, so the height inequality they established can be viewed as a generalization of this theorem to arbitrary projective varieties.

In this paper, we provide a simpler proof of the above results. Furthermore, we extend the results from the points of a projective variety to subschemes. The generalized result in terms of subschemes connects, as well as gives a clearer explanation to, the above mentioned result of McKinnon and Roth with the recent Diophantine approximation results in terms of the divisors obtained in [Corvaja and Zannier 2004; Evertse and Ferretti 2008; Levin 2014; Ru and Vojta 2016; Ru 2017].

We now state our result. Let V be a projective variety defined over a number field k.

MSC2010: primary 11J97; secondary 11J87, 14G05.
Keywords: Schmidt's subspace theorem, Roth's theorem, Diophantine approximation, Vojta's conjecture.

Definition 1.1. Let L be a line bundle over V with $h^0(V, NL) \geq 1$ for N big enough. Let Y be a proper closed subscheme of V and $\pi : \widetilde{V} \to V$ be the blow-up along Y, and E be the exceptional divisor. We define

$$\beta_{L,Y} := \liminf_{N\to\infty} \frac{\sum_{m=1}^{\infty} h^0(\widetilde{V}, N\pi^*L - mE)}{N \cdot h^0(V, NL)}.$$

Remark 1.2. (a) If Y is an effective Cartier divisor, then the blow-up is an isomorphism. Without loss of generality, we let π be the identity map, $\widetilde{V} = V$ and $E = Y$.

(b) Let D be an effective divisor on V, we define $\beta_{D,Y} := \beta_{\mathcal{O}(D),Y}$, where $\mathcal{O}(D)$ is the line sheaf associated to D.

(c) In the case when L is big, the $\lim_{N\to\infty}$ in the definition above exists. Indeed (see [McKinnon and Roth 2015, pp. 544–545]), we have

$$\beta_{L,Y} = \int_0^{\gamma_{\mathrm{eff}}} \frac{\mathrm{Vol}(L_\gamma)}{\mathrm{Vol}(L)} d\gamma,$$

where $L_\gamma := \pi^*L - \gamma E$ and $\gamma_{\mathrm{eff}} = \sup\{\gamma \geq 0 \mid L_\gamma \text{ is effective}\}$.

Definition 1.3. We say that the closed subschemes $Y_1, \ldots, Y_q$ of a projective variety V are in ℓ-subgeneral position if, for any $x \in V$, there are at most ℓ subschemes among $Y_1, \ldots, Y_q$ which contain x.

Remark 1.4. In the case that $Y_1 = y_1, \ldots, Y_q = y_q$ are points (as in [McKinnon and Roth 2015]), the condition that $y_1, \ldots, y_q$ *are distinct* implies that $Y_1, \ldots, Y_q$ *are in* 1*-subgeneral position* (i.e., with $\ell = 1$).

We establish the following result.

Main Theorem. *Let k be a number field and M_k be the set of places on k. Let $S \subset M_k$ be a finite subset containing all archimedean places. Let V be a projective variety defined over k and $Y_1, \ldots, Y_q$ be closed subschemes of V defined over k in ℓ-subgeneral position. For any $v \in S$, choose a local Weil function $\lambda_{Y_j,v}$ for each Y_j, $1 \leq j \leq q$. Let L be a big line bundle. Then for any $\epsilon > 0$*

$$\sum_{v\in S} \sum_{i=1}^{q} \lambda_{Y_i,v}(x) \leq \ell(\max_{1\leq i\leq q} \{\beta_{L,Y_i}^{-1}\} + \epsilon) h_L(x) \tag{1-1}$$

holds for all x outside a proper Zariski-closed subset Z of $V(k)$.

The following corollary of our main theorem recovers the main result of [McKinnon and Roth 2015]. The proof will be given in Section 3.

Corollary 1.5 [McKinnon and Roth 2015, Theorem 6.1]. *Let V be a projective variety over k. Then for any ample line bundle L and any $x \in V(\bar{k})$ either*

- $\alpha_x(L) \geq \beta_{L,x}$ *or*

- *there exists a proper subvariety $Z \subset V$, irreducible over $\bar{k}$, with $x \in Z(\bar{k})$ so that $\alpha_{x,V}(L) = \alpha_{x,Z}(L|Z)$, i.e., "$\alpha_x(L)$ is computed on a proper subvariety of V",*

where $\alpha_x(L)$ is the approximation constant defined in [McKinnon and Roth 2015, Definition 2.9], *and $\beta_{L,x}$ is defined in Definition 1.1* (*with Y taken as a point x*).

We will show in Lemma 2.2 that for any line bundle L, $x \in V$

$$\beta_{L,x} \geq \frac{n}{n+1}\epsilon_x(L), \tag{1-2}$$

where $n = \dim V$. We note that the Seshadri constant $\epsilon_x(L)$ does not decrease when restricting to a subvariety [McKinnon and Roth 2015, Proposition 3.4], so we can use induction to further get, from Corollary 1.5 and (1-2), the following result.

Corollary 1.6 [McKinnon and Roth 2015, Theorem 6.2, alternative statement]. *Let V be a projective variety over k. Let L be any ample line bundle and choose any $x \in V(\bar{k})$. Then for any $\delta > 0$, there are only finitely many solutions $y \in V(k)$ to*

$$d_v(x, y) < H_L(y)^{-((n+1)/(n\epsilon_x(L))+\delta)}.$$

In the case when $V = \mathbb{P}^n$ and $L = \mathcal{O}_{\mathbb{P}^n}(1)$, we have $\epsilon_x(L) = 1$ for all $x \in \mathbb{P}^n$ (see [McKinnon and Roth 2015, Lemma 3.3]). Therefore the above result generalizes the theorem of Roth.

We now turn to another extreme case when the subschemes $Y_1, \dots, Y_q$ are effective Cartier divisors $D_1, \dots, D_q$. Let $D := D_1 + \cdots + D_q$. Assume that each D_j is linearly equivalent to a fixed ample divisor A. Then we have the following relation of height functions $h_D = qh_A + O(1)$. On the other hand, by the Riemann–Roch theorem, with $n = \dim V$,

$$h^0(ND) = h^0(qNA) = \frac{(qN)^n A^n}{n!} + o(N^n)$$

and

$$h^0(ND - mD_j) = h^0((qN - m)A) = \frac{(qN - m)^n A^n}{n!} + o(N^n).$$

Thus

$$\sum_{m\geq 1} h^0(ND - mD_j) = \frac{A^n}{n!}\sum_{l=0}^{qN-1} l^n + o(N^{n+1}) = \frac{A^n(qN-1)^{n+1}}{(n+1)!} + o(N^{n+1}).$$

Hence

$$\beta_{D,D_j} = \lim_{N\to\infty} \frac{\frac{A^n(qN-1)^{n+1}}{(n+1)!} + o(N^{n+1})}{N\frac{(qN)^n A^n}{n!} + o(N^{n+1})} = \frac{q}{n+1}.$$

Thus the Main Theorem, together with the above computation, implies the following result of Chen, Ru, and Yan [2012] (see also [Corvaja and Zannier 2006]).

Theorem 1.7. *Let k be a number field and M_k the set of places on k. Let $S \subset M_k$ be a finite subset containing all archimedean places. Let V be a projective variety of dimension n defined over k. Let $D_1, \ldots, D_q$ be effective Cartier divisors in ℓ-subgeneral position on V. Assume that each D_j, $1 \le j \le q$, is linearly equivalent to a fixed ample divisor A. For any $v \in S$, choose a Weil function $\lambda_{D_j,v}$ for each D_j, $1 \le j \le q$. Then for any $\epsilon > 0$*

$$\sum_{v \in S} \sum_{i=1}^{q} \lambda_{D_i,v}(x) \le \ell(n+1+\epsilon) h_A(x) \tag{1-3}$$

holds for all x outside a proper Zariski-closed subset Z of $V(k)$. In particular, if $D_1, \ldots, D_q$ are in general position on V, then the inequality

$$\sum_{v \in S} \sum_{i=1}^{q} \lambda_{D_i,v}(x) \le n(n+1+\epsilon) h_A(x) \tag{1-4}$$

holds for all but finitely many $x \in V(k)$.

In the general case when $D_1, \ldots, D_q$ are only assumed to be big and nef, we can also compute β_{D,D_j}. The details will be carried out in the next section.

We note that recently the first named author and P. Vojta [2016] obtained the following sharp result in the case when $D_1, \ldots, D_q$ are in general position and when V is Cohen–Macaulay (for example when V is smooth).

Theorem 1.8 [Ru and Vojta 2016]. *Let k be a number field and M_k be the set of places on k. Let $S \subset M_k$ be a finite subset containing all archimedean places. Let V be a projective variety defined over k. Assume that V is Cohen–Macaulay. Let $D_1, \ldots, D_q$ be effective Cartier divisors in general position on V. For any $v \in S$, choose a Weil function $\lambda_{D_j,v}$ for each D_j, $1 \le j \le q$. Let L be a line bundle on V with $h^0(V, NL) \ge 1$ for N big enough. Then for any $\epsilon > 0$*

$$\sum_{v \in S} \sum_{i=1}^{q} \lambda_{D_i,v}(x) \le \left(\max_{1 \le i \le q} \{\beta_{L,D_i}^{-1}\} + \epsilon\right) h_L(x) \tag{1-5}$$

holds for all x outside a proper Zariski-closed subset Z of $V(k)$.

Theorem 1.8, together with the above computation, recovers the result of [Evertse and Ferretti 2002; 2008] in the case when V is smooth.

2. Computation of the constant $\beta_{L,Y}$

We first compute the constant $\beta_{L,y}$, i.e., we let $Y = y$ be a point in $V(k)$. The following lemma is a reformulation of Lemma 4.1 in [McKinnon and Roth 2015].

Lemma 2.1. *Let V be a projective variety and x be a point in V. Let $\pi : \widetilde{V} \to V$ be the blow-up along x, and E be the exceptional divisor. Let L be an ample line bundle and m a positive integer. Then*

(i) $h^0(\widetilde{V}, N\pi^*L - mE) = 0$ *if* $m > N \cdot \gamma_{\mathrm{eff},x}$, *where* $\gamma_{\mathrm{eff},x}$ *is defined in* [McKinnon and Roth 2015], *and*

(ii) $h^0(\widetilde{V}, N\pi^*L - mE) \geq h^0(V, NL) - m^n \operatorname{mult}_x V/n! + O(N^{n-1})$ *for* $N \gg 0$.

Proof. Write $h^0(\widetilde{V}, N\pi^*L - mE) = h^0(\widetilde{V}, N\pi^*L - N \cdot \gamma E)$, where $\gamma = m/N$. The argument in [McKinnon and Roth 2015] shows that $h^0(\widetilde{V}, N\pi^*L - mE) \geq h^0(V, NL) - m^n \operatorname{mult}_x V/n! + O(N^{n-1})$. □

The following is a restatement of Corollary 4.2 in [McKinnon and Roth 2015].

Lemma 2.2. *For any ample line bundle L, $x \in V$ and positive integer m, we have*

$$\beta_{L,x} \geq \frac{n}{n+1}\left(\frac{L^n}{\operatorname{mult}_x V}\right)^{1/n} \geq \frac{n}{n+1}\epsilon_x(L).$$

Proof. Choose a sufficiently large N. By Lemma 2.1 and the Riemann–Roch theorem,

$$h^0(\widetilde{V}, \pi^*NL - mE) \geq h^0(V, NL)\left(1 - \frac{\operatorname{mult}_x V}{L^n}\left(\frac{m}{N}\right)^n\right) + O(N^{n-1}). \tag{2-1}$$

The right-hand side of (2-1) is less than zero when $m > u = [N(L^n/\operatorname{mult}_x V)^{1/n}]$, hence

$$\sum_{m=1}^{\infty} h^0(\widetilde{V}, \pi^*NL - mE) \geq h^0(V, NL) \sum_{m=1}^{u}\left(1 - \frac{\operatorname{mult}_x V}{L^n}\left(\frac{m}{N}\right)^n\right) + O(N^n). \tag{2-2}$$

Consequently,

$$\begin{aligned}\beta_{L,x} &\geq \frac{1}{N}\sum_{m=1}^{u}\left(1 - \frac{\operatorname{mult}_x V}{L^n}\left(\frac{m}{N}\right)^n\right) + O\left(\frac{1}{N}\right)\\ &= \frac{1}{N}\left(u - \frac{\operatorname{mult}_x V}{L^n} \cdot \frac{u^{n+1}}{(n+1)N^n}\right) + O\left(\frac{1}{N}\right)\\ &\geq \frac{nu}{(n+1)N} + O\left(\frac{1}{N}\right).\end{aligned} \tag{2-3}$$

Let N run through all sufficiently large integers. Then we have

$$\beta_{L,x} \geq \frac{n}{n+1}\left(\frac{L^n}{\operatorname{mult}_x V}\right)^{1/n}.$$ □

Next we consider the case when $Y_j = D_j$, $1 \leq j \leq q$, are effective big and nef Cartier divisors on V.

Definition 2.3. Suppose X is a complete variety of dimension n. Let $D_1, \dots, D_q$ be effective Cartier divisors on X and let $D = D_1 + D_2 + \cdots + D_q$. We say that *$D$ has equidegree with respect to $D_1, D_2, \dots, D_q$* if $D_i \cdot D^{n-1} = D^n/q$ for all $i = 1, \dots, q$.

Lemma 2.4 [Levin 2009, Lemma 9.7]. *Let V be a projective variety of dimension n. If D_j, $1 \le j \le q$, are big and nef Cartier divisors on V, then there exist positive real numbers r_j such that $D = \sum_{j=1}^{q} r_j D_j$ has equidegree.*

Since divisors $r_j D_j$ and D_j have the same support, the above lemma tells us that we can always make the given big and nef divisors have equidegree without changing their supports. So now we assume that $D := D_1 + \cdots + D_q$ is of equidegree. To compute β_{D,D_j} for $j = 1, \dots, q$, we use the following lemma.

Lemma 2.5 [Autissier 2009, Lemma 4.2]. *Suppose E is a big and base-point free Cartier divisor on a projective variety V and F is a nef Cartier divisor on V such that $F - E$ is also nef. Let $\delta > 0$ be a positive real number. Then, for any positive integers N and m with $1 \le m \le \delta N$, we have*

$$h^0(NF - mE) \\ \ge \frac{F^n}{n!}N^n - \frac{F^{n-1}\cdot E}{(n-1)!}N^{n-1}m + \frac{(n-1)F^{n-2}\cdot E^2}{n!}N^{n-2}\min\{m^2, N^2\} + O(N^{n-1}),$$

where the implicit constant depends on β.

We compute $\sum_{m\ge 1} h^0(ND - mD_i)$ for each $1 \le i \le q$. Let $n = \dim V$ and assume that $n \ge 2$. Let $b = D^n/(nD^{n-1}\cdot D_i)$ and $A = (n-1)D^{n-2}\cdot D_i^2$. Then, by Lemma 2.5,

$$\begin{aligned}
\sum_{m=1}^{\infty} h^0(ND - mD_i) \\
&\ge \sum_{m=1}^{[bN]}\left(\frac{D^n}{n!}N^n - \frac{D^{n-1}\cdot D_i}{(n-1)!}N^{n-1}m + \frac{A}{n!}N^{n-2}\min\{m^2, N^2\}\right) + O(N^n) \\
&\ge \left(\frac{D^n}{n!}b - \frac{D^{n-1}\cdot D_i}{(n-1)!}\frac{b^2}{2} + \frac{A}{n!}g(b)\right)N^{n+1} + O(N^n) \\
&= \left(\frac{b}{2} + \frac{A}{D^n}g(b)\right)D^n\frac{N^{n+1}}{n!} + O(N^n) \\
&= \left(\frac{b}{2} + \alpha\right)Nh^0(ND) + O(N^n),
\end{aligned}$$

where $\alpha := g(b)A/D^n$ and $g : \mathbb{R}^+ \to \mathbb{R}^+$ is the function given by $g(x) = x^3/3$ if $x \le 1$ and $g(x) = x - \frac{2}{3}$ for $x \ge 1$. Now from the assumption of equidegree $D_i \cdot D^{n-1} = D^n/q$, so $b = q/n$. Moreover, $\alpha > 0$ since $\dim V \ge 2$ and the D_i are big and nef divisors. Hence

$$\beta_{D,D_i} = \liminf_N \frac{\sum_{m\geq 1} h^0(ND - mD_i)}{Nh^0(ND)} \geq \frac{b}{2} + \alpha.$$

Thus we have proved the following.

Proposition 2.6. *Let V be a projective variety of* $\dim V \geq 2$ *and assume that $D := \sum_{j=1}^q D_j$ has equidegree with respect to $D_1, \ldots, D_q$ which are big and nef. Then*

$$\beta_{D,D_i} = \liminf_N \frac{\sum_{m\geq 1} h^0(ND - mD_i)}{Nh^0(ND)} > \frac{q}{2n} + \alpha,$$

where α is a computable positive number.

Proposition 2.6, together with the Main Theorem, implies the following result.

Theorem 2.7 [Hussein and Ru 2018]. *Let k be a number field and let $S \subseteq M_k$ be a finite set containing all archimedean places. Let V be a projective variety of dimension ≥ 2 over k and let $D_1, \ldots, D_q$ be effective, big, and nef Cartier divisors on V defined over k, located in ℓ-subgeneral position. Let $r_i > 0$ be real numbers such that $D := \sum_{i=1}^q r_i D_i$ has equidegree* (*such numbers exist due to Lemma 2.4*). *Then, for $\epsilon_0 > 0$ small enough, the inequality*

$$\sum_{v\in S}\sum_{j=1}^q r_j \lambda_{D_i,v}(x) < \ell\left(\frac{2\dim V}{q} - \epsilon_0\right)\left(\sum_{j=1}^q r_j h_{D_j}(x)\right)$$

holds for all $x \in V(k)$ outside a proper Zariski-closed subset of V.

3. Proof of the Main Theorem

We first recall some basic properties of local Weil functions associated to closed subschemes from [Silverman 1987, Section 2]. We assume that the readers are familiar with the notion of Weil functions associated to divisors (see [Lang 1983, Chapter 10], [Hindry and Silverman 2000, B.8] or [Silverman 1987, Section 1]).

Let Y be a closed subscheme on a projective variety V defined over k. Then one can associate to each place $v \in M_k$ a function

$$\lambda_{Y,v} : V \setminus \mathrm{supp}(Y) \to \mathbb{R}$$

satisfying some functorial properties (up to an M_k-constant) described in [Silverman 1987, Theorem 2.1]. Intuitively, for each $P \in V$ and $v \in M_k$,

$$\lambda_{Y,v}(P) = -\log(v\text{-adic distance from } P \text{ to } Y).$$

The following lemma indicates the existence of local Weil functions.

Lemma 3.1. *Let Y be a closed subscheme of V. There exist effective divisors $D_1, \dots, D_r$ such that*

$$Y = \cap D_i.$$

Proof. See Lemma 2.2 from [Silverman 1987]. □

Definition 3.2. Let k be a number field, and M_k be the set of places on k. Let V be a projective variety over k and let $Y \subset V$ be a closed subscheme of V. We define the (local) Weil function for Y with respect to $v \in M_k$ as

$$\lambda_{Y,v} = \min_i \{\lambda_{D_i,v}\}, \tag{3-1}$$

when $Y = \cap D_i$ (such D_i exist according to the above lemma).

Lemma 3.3 [Vojta 1987, Lemma 2.5.2; Silverman 1987, Theorem 2.1(h)]. *Let Y be a closed subscheme of V and let $\widetilde{V}$ be a blow-up of V along Y with exceptional divisor $E = \pi^*Y$. Then $\lambda_{Y,v}(\pi(P)) = \lambda_{E,v}(P) + O_v(1)$ for $P \in \widetilde{V}$.*

Note that in the original statement of Lemma 2.5.2 in [Vojta 1987], V is assumed to be smooth, but from the proof it is easy to see that it works for a general projective variety from Theorem 2.1(h) in [Silverman 1987].

For our purpose, it suffices to fix a choice of local Weil functions $\lambda_{Y_i,v}$ for each $1 \le i \le q$ and $v \in S$.

Lemma 3.4. *Let $Y_1, \dots, Y_q$ be closed subschemes of a projective variety V in ℓ-subgeneral position. Then*

$$\sum_{i=1}^{q} \lambda_{Y_i,v}(x) \le \max_I \sum_{j\in I} \lambda_{Y_j,v}(x) + O_v(1), \tag{3-2}$$

where I runs over all index subsets of $\{1, \dots, q\}$ with ℓ elements for all $x \in V(k)$.

Proof. Let $\{i_1, \dots, i_q\} = \{1, \dots, q\}$. Since the Y_i, $1 \le i \le q$, are in ℓ-subgeneral position, $\bigcap_{t=1}^{\ell+1} Y_{i_t} = \varnothing$. Then

$$\min_{1\le i\le \ell+1} \{\lambda_{Y_i,v}\} = \{\lambda_{\bigcap_{t=1}^{\ell+1} Y_{i_t},v}\} = O_v(1). \tag{3-3}$$

We note that the first equality follows from (3-1), the definition of the local Weil function; and the second equality follows from Corollary 3.3 in [Lang 1983, Chapter 10]. For x with the following ordering

$$\lambda_{Y_{i_1},v}(x) \ge \lambda_{Y_{i_2},v}(x) \ge \cdots \ge \lambda_{Y_{i_q},v}(x),$$

we have

$$\sum_{i=1}^{q} \lambda_{Y_i,v}(x) = \sum_{i=1}^{\ell} \lambda_{Y_i,v}(x) + O_v(1).$$

Then assertion (3-2) follows directly as the number of subvarieties under consideration is finite. □

We also need the following generalized Schmidt subspace theorem.

Theorem 3.5 [Ru and Vojta 2016, Theorem 2.7]. *Let k be a number field, S be a finite set of places of k containing all archimedean places, X be a complete variety over k, D be a Cartier divisor on X, W be a nonzero linear subspace of $H^0(X, \mathcal{O}(D))$, $s_1, \dots, s_q$ be nonzero elements of W, $\epsilon > 0$, and $c \in \mathbb{R}$. For each $j = 1, \dots, q$, let D_j be the Cartier divisor (s_j) and λ_{D_j} be a Weil function for D_j. Then there is a proper Zariski-closed subset Z of X, depending only on k, S, X, D, W, $s_1, \dots, s_q$, ϵ, c, and the choices of Weil and height functions, such that the inequality*

$$\sum_{v\in S} \max_J \sum_{j\in J} \lambda_{D_j,v}(x) \le (\dim W + \epsilon) h_D(x) + c \tag{3-4}$$

holds for all $x \in (X \setminus Z)(k)$. Here the set J ranges over all subsets of $\{1, \dots, q\}$ such that the sections $(s_j)_{j\in J}$ are linearly independent.

We are now ready to prove the Main Theorem.

Proof of the Main Theorem. Let $\delta > 0$ be a sufficiently small number. We may choose a sufficiently large integer N such that, for $i = 1, \dots, q$,

$$\sum_{m=1}^{\infty} h^0(\widetilde{V}_i, N\pi^*L - E_i) \ge (\beta_{L,Y_i} - \delta) N h^0(V, NL), \tag{3-5}$$

where $\pi_i : \widetilde{V}_i \to V$ is the blow-up at Y_i and $E_i = \pi^{-1}(Y_i)$ is he exceptional divisor of π_i.

Let $x \in V(k)$ and $v \in S$. Since the Y_i, $1 \le i \le q$, are in ℓ-subgeneral position, it follows from Lemma 3.4 that

$$\sum_{i=1}^{q} \lambda_{Y_i,v}(x) \le \ell \lambda_{Y_{i_0},v}(x) + O_v(1), \tag{3-6}$$

for some i_0 with $1 \le i_0 \le q$, where the constant $O_v(1)$ is independent of x. Note that i_0 depends on the point x, but $O_v(1)$ is independent of x.

Write $\widetilde{V}_{i_0}$ as $\widetilde{V}$, π_{i_0} as π and E_{i_0} as E. We consider the following filtration.

$$H^0(\widetilde{V}, \pi^*NL) \supseteq H^0(\widetilde{V}, \pi^*NL - E) \supseteq H^0(\widetilde{V}, \pi^*NL - 2E) \supseteq \cdots \tag{3-7}$$

We identify $H^0(V, NL)$ with $H^0(\widetilde{V}, \pi^*NL)$ as vector spaces (note: according to the footnote on page 553 in [McKinnon and Roth 2015], if X is not normal, then $H^0(V, NL)$ may only be a proper subspace of $H^0(\widetilde{V}, \pi^*NL)$. However, since the volume is a birational constant, the asymptotic calculations go through without change). Choose regular sections $s_1, \dots, s_M \in H^0(V, NL)$ successively so that their pull-back $\pi^*s_1, \dots, \pi^*s_M \in H^0(\widetilde{V}, \pi^*NL)$ form a basis associated to this filtration, where $M = h^0(\widetilde{V}, N\pi^*L)$. For a section $\pi^*s \in H^0(\widetilde{V}, \pi^*NL - mE)$ (regarded as a subspace of $H^0(\widetilde{V}, \pi^*NL)$) we have

$$\operatorname{div}(\pi^*s) \geq mE. \tag{3-8}$$

Hence, $\lambda_{(\pi^*s),v} \geq m\lambda_{E,v} + O_v(1)$. Note that although $O_v(1)$ here depends on i_0 (which depends on x), there are q choices of such i_0 and V is compact, so we can again make $O_v(1)$ independent of x. Therefore, also using Lemma 3.3 and (3-5),

$$\begin{aligned}
&\sum_{j=1}^{M} \lambda_{(\pi^*s_j),v} \\
&\quad \geq \sum_{m=1}^{\infty} m(h^0(\widetilde{V}, \pi^*NL - mE) - h^0(\widetilde{V}, \pi^*NL - (m+1)E))\lambda_{E,v} + O_v(1) \\
&\quad = \sum_{m=1}^{\infty} m(h^0(\widetilde{V}, \pi^*NL - mE) - h^0(\widetilde{V}, \pi^*NL - (m+1)E))\lambda_{Y_{i_0},v} \circ \pi + O_v(1) \\
&\quad = \sum_{m=1}^{\infty} h^0(\widetilde{V}, \pi^*NL - mE)\lambda_{Y_{i_0},v} \circ \pi + O_v(1) \\
&\quad \geq (\beta_{L,Y_{i_0}} - \delta)Nh^0(V, NL)\lambda_{Y_{i_0},v} \circ \pi + O_v(1).
\end{aligned}$$

The functorial property of Weil functions implies $\lambda_{(\pi^*s_j),v} = \lambda_{(s_j),v} \circ \pi + O_v(1)$. Hence, the above inequality, together with (3-6), implies that

$$\begin{aligned}
&\sum_{i=1}^{q} \lambda_{Y_i,v}(x) \\
&\quad \leq \frac{\ell}{N \cdot h^0(V, NL)(\min_{1 \leq i \leq q}\{\beta_{L,Y_i}\} - \delta)} \max_J \left\{ \sum_{j \in J} \lambda_{(s_j),v}(x) \right\} + O_v(1),
\end{aligned} \tag{3-9}$$

where J is a subset containing M linearly independent sections taken among the collection of sections $\{s_j(i_0, v) | 1 \leq i_0 \leq q, v \in S\}$ coming from the claim (3-6). It then follows from Theorem 3.5 and a suitable choice of δ that for a given $\epsilon > 0$ there exists a proper algebraic subset Z of V defined over k such that

$$\sum_{v \in S} \sum_{i=1}^{q} \lambda_{Y_i,v}(x) \leq (\ell \cdot \max_{1 \leq i \leq q} \{\beta_{L,Y_i}^{-1}\} + \epsilon)h_L(x), \tag{3-10}$$

for all $x \in V(k) \setminus Z(k)$. □

Proof of Corollary 1.5. Let v be a place of k. The main point of the proof is to reformulate the distance function $d_v(\cdot, \cdot)$ defined on $V(\bar{k})$ [McKinnon and Roth 2015, Section 2] into a product of several distance functions on $V(K)$, where K is a finite extension of k. Following the construction in [McKinnon and Roth 2015, Section 2], we fix an extension of v to $\bar{k}$. The place defines an absolute value $\|\cdot\|_v$ on $\bar{k}$. If $K \subset \bar{k}$ is a finite extension of k, then $d_v(\,\cdot\,,\,\cdot\,)_K = d_v(\,\cdot\,,\,\cdot\,)_k^{[K_v:k_v]}$. Here $d_v(\,\cdot\,,\,\cdot\,)_K$ refers to the distance function defined by using the same embedding and normalizing with respect to K and $d_v(\,\cdot\,,\,\cdot\,)_k$ the distance function normalized with respect to k (see [McKinnon and Roth 2015, Proposition 2.1(b)]). Assume that $V \subset \mathbb{P}^N$ (given by the canonical map associated to the ample line bundle L). For a given fixed point $y = [y_0 : \cdots : y_N] \in V(\bar{k})$, let K be the Galois closure of $k(y_0, \ldots, y_N)$ over k. For each $v \in M_k$, the inclusion map $(i_v)|_K : K \to \bar{k}_v$ induces a place $w_0 := v$ of K over v, and other places w of K over v are conjugates by elements $\sigma_w \in \mathrm{Gal}(K/k)$ such that $\|\sigma_w(a)\|_w = \|a\|_v$ for all $a \in K$. Then, for $x, y \in K$,

$$\begin{aligned}\prod_{w \in M_K, w|v} d_w(\sigma_w(x), \sigma_w(y))_K &= \prod_{w \in M_K, w|v} d_v(x, y)_K \\ &= \prod_{w \in M_K, w|v} d_v(x, y)_k^{[K_v:k_v]} \\ &= [K:k] d_v(x, y)_k,\end{aligned}$$

i.e.,

$$d_v(x, y)_k = \prod_{w \in M_K, w|v} d_w(\sigma_w(x), \sigma_w(y))_K^{1/[K:k]}, \quad \text{for } x, y \in K. \tag{3-11}$$

To compute $\alpha_y(L)$, we consider any sequence $\{x_i\} \subseteq V(k)$ of distinct points with $d_v(y, x_i) \to 0$. By (3-11), we have $d_v(y, x_i)_k = \prod_{w \in M_K, w|v} d_w(\sigma_w(y), x_i)_K^{1/[K:k]}$. (Here we extend $\sigma_w \in \mathrm{Gal}(K/k)$ to the map from $V(K)$ to $V(K)$ by acting on the coordinates of the points.) The distance function $d_w(y, x)$ in [McKinnon and Roth 2015] is constructed by choosing an embedding $\phi_L : V \to \mathbb{P}^N$ into a projective space via the sections of L and measuring the distance in the embedded space. For a fixed y we denote $-\log d_w(y, \cdot\,)$ by $\lambda_{\phi(y),w}$, which is a local Weil function on the embedded space. We note that this fact can also be proved by a slight modification of Lemma 2.6 in [McKinnon and Roth 2015]. By the functoriality of Weil functions of closed subschemes [Silverman 1987, Theorem 2.1(h)] we have $-\log d_w(\sigma_w(y), x) = \lambda_{\sigma_w(y),w}(x) + O(1)$. On the other hand, it is clear from the definition that $\beta_{L,y} = \beta_{L,\sigma_w(y)}$ for very $\sigma_w \in \mathrm{Gal}(K/k)$. The Main Theorem then implies that for any $\epsilon > 0$

$$\log d_v(y, x_i) = \frac{1}{[K:k]} \sum_{w \in M_K, w|v} \log d_w(y, x_i) \geq -(\{\beta_{L,y}^{-1}\} + \epsilon) h_L(x_i) \tag{3-12}$$

holds for all x_i outside a proper Zariski-closed subset Z of $V(K)$ (note that, in this case, $\ell = 1$). We note that Z is indeed defined over k since all the x_i are in k. In conclusion, we have shown that for all sequences $\{x_i\} \subseteq V(k)$ of distinct points with $d_v(y, x_i) \to 0$, if $\alpha_y(\{x_i\}, L) < \beta_{L,y}$, then all but finitely many of the points of $\{x_i\}$ lie in Z. If (a) holds, then we are done. Therefore we assume that $\alpha_y(L) > \beta_{L,y}$.

Then the previous conclusion shows, in this case, that $\alpha_y(L) = \alpha_{y,Z}(L|_Z)$. To see Z is irreducible over $\bar{k}$, we first use Proposition 2.14(f) in [McKinnon and Roth 2015] to reduce Z to one of the irreducible components of Z over k, say Y such that $\alpha_{y,Z}(L|_Z) = \alpha_{y,Y}(L|_Y)$. Without loss of generality we can assume that $Z = Y$, i.e., Z itself is irreducible over k. We then apply Lemma 2.17 in [McKinnon and Roth 2015] to conclude that Z is indeed geometrically irreducible, i.e., Z is irreducible over $\bar{k}$. $\square$

4. The complex case

In this section, we consider the analogous result of our Main Theorem in Nevanlinna theory. Let V be a complex projective variety. We use the standard notation in Nevanlinna theory (see, for example, [Ru 2016]). Note that the Weil function for divisors has been defined, so the Weil function λ_Y for a subscheme $Y \subset V$ can also be defined using Lemma 3.1, similar to Definition 3.2. We define, for a holomorphic map $f : \mathbb{C} \to V$ with $f(\mathbb{C}) \not\subset Y$, the proximity function

$$m_f(r, Y) = \int_0^{2\pi} \lambda_Y(f(re^{i\theta}))\,\frac{d\theta}{2\pi}.$$

We note that all the properties used above about the Weil functions in the arithmetic case hold for the complex case (see, for example, [Ru 2016; Ru and Vojta 2016]).

Theorem 4.1. *Let V be a complex projective variety and $Y_1, \dots, Y_q$ be closed subschemes of V in ℓ-subgeneral position. Let L be a big line bundle. Let $f : \mathbb{C} \to V$ be a holomorphic map with Zariski dense image. Then for any $\epsilon > 0$*

$$\sum_{i=1}^{q} m_f(r, Y_i) \le \ell\big(\max_{1\le i\le q}\{\beta_{L,Y_i}^{-1}\} + \epsilon\big) T_{f,L}(r)\,\|, \tag{4-1}$$

where $\|$ means that the inequality holds for all $r \in (0, +\infty)$ outside a set of finite Lebesgue measure.

To prove the theorem, we need the following result.

Theorem 4.2 [Ru and Vojta 2016, Theorem 2.8]. *Let X be a complex projective variety, D be a Cartier divisor on X, W be a nonzero linear subspace of $H^0(X, \mathcal{O}(D))$, and $s_1, \dots, s_q$ be nonzero elements of W. Let $f : \mathbb{C} \to X$ be a holomorphic map with Zariski-dense image. Then*

$$\int_0^{2\pi} \max_J \sum_{j\in J} \lambda_{(s_j)}(f(re^{i\theta}))\,\frac{d\theta}{2\pi} \le (\dim W) T_{f,D}(r) + O(\log^+ T_{f,D}(r)) + o(\log r)\,\|,$$

where the set J ranges over all subsets of $\{1, \dots, q\}$ such that the sections $(s_j)_{j\in J}$ are linearly independent.

Proof of Theorem 4.1. Similar to the proof of the Main Theorem, let $\delta > 0$ be a sufficiently small number. We choose N large enough that, for $i = 1, \ldots, q$,

$$\sum_{m=1}^{\infty} h^0(\widetilde{V}_i, N\pi_i^* L - mE_i) \geq (\beta_{L,Y_i} - \delta) N h^0(V, NL).$$

Let $x \in V$. Since Y_i, $1 \leq i \leq q$, are in ℓ-subgeneral position, similar to Lemma 3.4, we have

$$\sum_{i=1}^{q} \lambda_{Y_i}(x) \leq \ell \lambda_{Y_{i_0}}(x) + O(1), \tag{4-2}$$

for some i_0 with $1 \leq i_0 \leq q$, where i_0 depends on the point x, but $O(1)$ is independent of x.

Let $\pi : \widetilde{V} \to V$ be the blow-up at Y_{i_0} and $E = \pi^{-1}(Y_{i_0})$ be the exceptional divisor of π. We consider the filtration of $H^0(\widetilde{V}, \pi^* NL)$ defined in (3-7). By identifying $H^0(V, NL)$ with $H^0(\widetilde{V}, \pi^* NL)$ as vector spaces, we can choose regular sections $s_1, \ldots, s_M \in H^0(V, NL)$, where $M = h^0(V, NL)$, successively so that their pullbacks $\pi^* s_1, \ldots, \pi^* s_M \in H^0(\widetilde{V}, \pi^* NL)$ form a basis associated to this filtration. Then, in the same way as deriving (3-9), we can get

$$\sum_{i=1}^{q} \lambda_{Y_i}(x) \leq \frac{\ell}{N \cdot h^0(V, NL)(\min_{1 \leq i \leq q}\{\beta_{L,Y_i}\} - \delta)} \sum_{j=1}^{q} \lambda_{(s_j)}(x) + O(1).$$

Note that the basis $\{s_1, \ldots, s_M\}$ depends only on i_0, so the number of such choices is finite, since $i_0 \in \{1, \ldots, q\}$, while x varies in (4-2). We denote the set of bases as $J_1, \ldots, J_T$. Thus we get, for every $x \in V$,

$$\sum_{i=1}^{q} \lambda_{Y_i}(x) \leq \frac{\ell}{N \cdot h^0(V, NL)(\min_{1 \leq i \leq q}\{\beta_{L,Y_i}\} - \delta)} \max_{1 \leq t \leq T} \sum_{j \in J_t} \lambda_{(s_j)}(x) + O(1).$$

By taking $x = f(re^{i\theta})$ and then integrating, it then follows from Theorem 4.2 and a suitable choice of δ that, for the given $\epsilon > 0$,

$$\sum_{i=1}^{q} \int_0^{2\pi} \lambda_{Y_i}(f(re^{i\theta})) \frac{d\theta}{2\pi} \leq \ell (\max_{1 \leq i \leq q} \{\beta_{\mathcal{L},Y_i}^{-1}\} + \epsilon) T_{f,\mathcal{L}}(r) \big\|.$$

This finishes the proof. □

Theorem 4.1, together with Lemma 2.2, implies the following corollary.

Corollary 4.3. *Let V be a complex projective variety of dimension n and $a_1, \ldots, a_q$ be distinct points on V. Let L be an ample line bundle. Let $f : \mathbb{C} \to V$ be a* *holomorphic map with Zariski dense image. Then for any* $\epsilon > 0$,

$$\sum_{i=1}^{q} m_f(r, a_i) \leq \left(\frac{n+1}{n} \max_{1\leq i\leq q} \{\epsilon_{a_i}^{-1}(L)\} + \epsilon\right) T_{f,L}(r) \|,$$

where $\epsilon_x(L)$ *is the Seshadri constant of* L *at the point* $x \in V$.

In particular, if $V = \mathbb{P}^n$, *then for any* $\epsilon > 0$,

$$\sum_{i=1}^{q} m_f(r, a_i) \leq \left(\frac{n+1}{n} + \epsilon\right) T_{f,L}(r) \|.$$

Acknowledgments

The first named author wishes to thank the Institute of Mathematics, Academia Sinica, Taiwan for kind hospitality during which part of the work on this paper took place. The authors thank Shin-Yao Jow for pointing out a mistake in an earlier version and helpful discussions. The authors also wish to thank the referee for many helpful suggestions.

References

[Autissier 2009] P. Autissier, "Géométries, points entiers et courbes entières", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:2 (2009), 221–239. MR Zbl

[Chen et al. 2012] Z. Chen, M. Ru, and Q. Yan, "The degenerated second main theorem and Schmidt's subspace theorem", *Sci. China Math.* **55**:7 (2012), 1367–1380. MR Zbl

[Corvaja and Zannier 2004] P. Corvaja and U. Zannier, "On a general Thue's equation", *Amer. J. Math.* **126**:5 (2004), 1033–1055. MR Zbl

[Corvaja and Zannier 2006] P. Corvaja and U. Zannier, "Addendum to 'On a general Thue's equation' ", *Amer. J. Math.* **128**:4 (2006), 1057–1066. MR Zbl

[Evertse and Ferretti 2002] J.-H. Evertse and R. G. Ferretti, "Diophantine inequalities on projective varieties", *Int. Math. Res. Not.* **2002**:25 (2002), 1295–1330. MR Zbl

[Evertse and Ferretti 2008] J.-H. Evertse and R. G. Ferretti, "A generalization of the Subspace Theorem with polynomials of higher degree", pp. 175–198 in *Diophantine approximation*, edited by H. P. Schlickewei et al., Dev. Math. **16**, Springer, 2008. MR Zbl

[Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics **201**, Springer, 2000. MR Zbl

[Hussein and Ru 2018] S. Hussein and M. Ru, "A general defect relation and height inequality for divisors in subgeneral position", preprint, 2018. To appear in *Asian Journal of Mathematics*.

[Lang 1983] S. Lang, *Fundamentals of Diophantine geometry*, Springer, 1983. MR Zbl

[Levin 2009] A. Levin, "Generalizations of Siegel's and Picard's theorems", *Ann. of Math.* (2) **170**:2 (2009), 609–655. MR Zbl

[Levin 2014] A. Levin, "On the Schmidt subspace theorem for algebraic points", *Duke Math. J.* **163**:15 (2014), 2841–2885. MR Zbl

[McKinnon and Roth 2015] D. McKinnon and M. Roth, "Seshadri constants, diophantine approximation, and Roth's theorem for arbitrary varieties", *Invent. Math.* **200**:2 (2015), 513–583. MR Zbl

[Ru 2016] M. Ru, "A defect relation for holomorphic curves intersecting general divisors in projective varieties", *J. Geom. Anal.* **26**:4 (2016), 2751–2776. MR Zbl

[Ru 2017] M. Ru, "A general Diophantine inequality", *Funct. Approx. Comment. Math.* **56**:2 (2017), 143–163.

[Ru and Vojta 2016] M. Ru and P. Vojta, "A birational Nevanlinna constant and its consequences", preprint, 2016. arXiv

[Silverman 1987] J. H. Silverman, "Arithmetic distance functions and height functions in Diophantine geometry", *Math. Ann.* **279**:2 (1987), 193–216. MR Zbl

[Vojta 1987] P. Vojta, *Diophantine approximations and value distribution theory*, Lecture Notes in Mathematics **1239**, Springer, 1987. MR Zbl

Communicated by Joseph H. Silverman
Received 2016-09-01 Revised 2017-02-06 Accepted 2017-03-31

minru@math.uh.edu *Department of Mathematics, East China Normal University, Shanghai, China*

Department of Mathematics, University of Houston, Houston, TX, United States

jwang@math.sinica.edu.tw *Institute of Mathematics, Academia Sinica, Taipei, Taiwan*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2339

Variation of anticyclotomic Iwasawa invariants in Hida families

Francesc Castella, Chan-Ho Kim and Matteo Longo

Building on the construction of big Heegner points in the quaternionic setting by Longo and Vigni, and their relation to special values of Rankin–Selberg L-functions established by Castella and Longo, we obtain anticyclotomic analogues of the results of Emerton, Pollack and Weston on the variation of Iwasawa invariants in Hida families. In particular, combined with the known cases of the anticyclotomic Iwasawa main conjecture in weight 2, our results yield a proof of the main conjecture for p-ordinary newforms of higher weights and trivial nebentypus.

Introduction

In the remarkable paper [Emerton et al. 2006], Emerton, Pollack and Weston obtained striking results on the behavior of the cyclotomic Iwasawa invariants attached to p-ordinary modular forms as they vary in Hida families. In particular, combined with Greenberg's conjecture on the vanishing of the μ-invariant, their main result *reduces* the proof of the main conjecture to the weight two case. In this paper, we develop analogous results for newforms base-changed to imaginary quadratic fields in the definite anticyclotomic setting. In particular, combined with

MSC2010: primary 11R23; secondary 11F33.
Keywords: Iwasawa theory, Hida theory, Selmer groups, Heegner points, special values of L-functions.

Vatsal's result [2003] on the vanishing of the anticyclotomic μ-invariant, and the known cases of the anticyclotomic main conjecture in weight 2 (thanks to the works of Bertolini and Darmon [2005], Pollack and Weston [2011], and Skinner and Urban [2014]), our results yield a proof of Iwasawa's main conjecture for p-ordinary modular forms of higher weights $k \geqslant 2$ and trivial nebentypus in the anticyclotomic setting.

Let us begin by recalling the setup of [Emerton et al. 2006], but adapted to the context at hand. Let

$$\bar{\rho} : G_{\mathbb{Q}} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F})$$

be a continuous Galois representation defined over a finite field $\mathbb{F}$ of characteristic $p > 3$, and assume that $\bar{\rho}$ is odd and irreducible. After the proof of Serre's conjecture [Khare and Wintenberger 2009], we know that $\bar{\rho}$ is modular, meaning that $\bar{\rho}$ is isomorphic to the mod p Galois representation $\bar{\rho}_{f_0}$ associated to an elliptic newform f_0. Throughout this paper, it will be assumed that $\bar{\rho} \simeq \bar{\rho}_{f_0}$ for some newform f_0 of weight 2 and trivial nebentypus.

Let $N(\bar{\rho})$ be the tame conductor of $\bar{\rho}$, and let $K/\mathbb{Q}$ be an imaginary quadratic field of discriminant prime $-D_K < 0$ to $pN(\bar{\rho})$. The field K then determines a decomposition

$$N(\bar{\rho}) = N(\bar{\rho})^+ \cdot N(\bar{\rho})^-$$

with $N(\bar{\rho})^+$ (resp. $N(\bar{\rho})^-$) only divisible by primes which are split (resp. inert) in K. We similarly define the decomposition $M = M^+ \cdot M^-$ for any positive integer M prime to D_K.

As in [Pollack and Weston 2011], we consider the following conditions on a pair $(\bar{\rho}, N^-)$, where N^- is a fixed square-free product of an odd number of primes inert in K:

Assumption (CR). (1) $\bar{\rho}$ is irreducible;

(2) $N(\bar{\rho})^- \mid N^-$;

(3) $\bar{\rho}$ is ramified at every prime $\ell \mid N^-$ such that $\ell \equiv \pm 1 \pmod{p}$.

Let $\mathcal{H}(\bar{\rho})$ be the set of all p-ordinary and p-stabilized newforms with mod p Galois representation isomorphic to $\bar{\rho}$, and let $\Gamma := \mathrm{Gal}(K_\infty/K)$ denote the Galois group of the anticyclotomic $\mathbb{Z}_p$-extension of K. Associated with each $f \in \mathcal{H}(\bar{\rho})$ of tame level N_f with $N_f^- = N^-$, defined over say a finite extension $F/\mathbb{Q}_p$ with ring of integers $\mathcal{O}$, there is a p-adic L-function

$$L_p(f/K) \in \mathcal{O}[\![\Gamma]\!]$$

constructed by Bertolini and Darmon [1996] in weight two, and by Chida and Hsieh [2016] for higher weights. The p-adic L-function $L_p(f/K)$ is characterized, as

χ runs over the p-adic characters of Γ corresponding to certain algebraic Hecke characters of K, by an interpolation property of the form

$$\chi(L_p(f/K)) = C_p(f,\chi)\cdot E_p(f,\chi)\cdot \frac{L(f/K,\chi,k/2)}{\Omega_{f,N^-}},$$

where $C_p(f,\chi)$ is an explicit nonzero constant, $E_p(f,\chi)$ is a p-adic multiplier, and Ω_{f,N^-} is a complex period making the above ratio algebraic. (Of course, implicit in all the above is a fixed choice of complex and p-adic embeddings $\mathbb{C} \overset{\iota_\infty}{\hookleftarrow} \overline{\mathbb{Q}} \overset{\iota_p}{\hookrightarrow} \overline{\mathbb{Q}}_p$.)

The anticyclotomic Iwasawa main conjecture gives an arithmetic interpretation of $L_p(f/K)$. More precisely, let

$$\rho_f : G_{\mathbb{Q}} \to \mathrm{Aut}_F(V_f) \simeq \mathrm{GL}_2(F)$$

be a self-dual twist of the p-adic Galois representation associated to f, fix an $\mathcal{O}$-stable lattice $T_f \subseteq V_f$, and set $A_f := V_f/T_f$. Let $D_p \subseteq G_{\mathbb{Q}}$ be the decomposition group corresponding to our fixed embedding ι_p, and let $\varepsilon_{\mathrm{cyc}}$ be the p-adic cyclotomic character. Since f is p-ordinary, there is a unique one-dimensional D_p-invariant subspace $F_p^+V_f \subseteq V_f$ where the inertia group at p acts via $\varepsilon_{\mathrm{cyc}}^{k/2}\psi$, with ψ a finite order character. Let $F_p^+A_f$ be the image of $F_p^+V_f$ in A_f and set $F_p^-A_f := A_f/F_p^+A_f$. Following the terminology in [Pollack and Weston 2011], the *minimal Selmer group* of f is defined by

$$\mathrm{Sel}(K_\infty, f) := \ker\left\{ H^1(K_\infty, A_f) \to \prod_{w\nmid p} H^1(K_{\infty,w}, A_f) \times \prod_{w\mid p} H^1(K_{\infty,w}, F_p^-A_f)\right\},$$

where w runs over the places of K_∞. By standard arguments (see [Greenberg 1989], for example), one knows that the Pontryagin dual of $\mathrm{Sel}(K_\infty, f)$ is finitely generated over the anticyclotomic Iwasawa algebra $\Lambda := \mathcal{O}[\![\Gamma]\!]$. The *anticyclotomic main conjecture* is then the following:

Conjecture 1. *The Pontryagin dual* $\mathrm{Sel}(K_\infty, f)^\vee$ *is* Λ*-torsion, and*

$$Ch_\Lambda(\mathrm{Sel}(K_\infty, f)^\vee) = (L_p(f/K)).$$

For newforms f of weight 2 corresponding to elliptic curves $E/\mathbb{Q}$ with ordinary reduction at p, and under rather stringent assumptions on $\bar\rho_f$ which were later relaxed by Pollack and Weston [2011], one of the divisibilities predicted by Conjecture 1 was obtained by Bertolini and Darmon [2005] using Heegner points and Kolyvagin's method of Euler systems. More recently, after the work of Chida and Hsieh [2015] the divisibility

$$Ch_\Lambda(\mathrm{Sel}(K_\infty, f)^\vee) \supseteq (L_p(f/K))$$

is known for newforms f of weight $k \leqslant p-2$ and trivial nebentypus, provided the pair $(\bar\rho_f, N_f^-)$ satisfies a mild strengthening of Hypotheses (CR). This restriction to small weights comes from the use of Ihara's lemma [Diamond and Taylor 1994], and it seems difficult to directly extend their arguments in [Chida and Hsieh 2015] to higher weights. Instead, as we shall explain in the following paragraphs, in this paper we will complete the proof of Conjecture 1 to all weights $k \equiv 2 \pmod{p-1}$ by a different approach, using Howard's big Heegner points in Hida families [Howard 2007], as extended by Longo and Vigni [2011] to quaternionic Shimura curves.

Associated with every $f \in \mathcal{H}(\bar{\rho})$ there are *anticyclotomic Iwasawa invariants* $\mu^{\mathrm{an}}(K_\infty, f)$, $\lambda^{\mathrm{an}}(K_\infty, f)$, $\mu^{\mathrm{alg}}(K_\infty, f)$, and $\lambda^{\mathrm{alg}}(K_\infty, f)$. The analytic (resp. algebraic) λ-invariants are the number of zeros of $L_p(f/K)$ (resp. of a generator of the characteristic ideal of $\mathrm{Sel}(K_\infty, f)^\vee$), while the μ-invariants are defined as the exponent of the highest power of ϖ (with $\varpi \in \mathcal{O}$ any uniformizer) dividing the same objects. Our main results on the variation of these invariants are summarized in the following. (Recall that we assume $\bar{\rho} \simeq \bar{\rho}_{f_0}$ for some newform f_0 of weight 2 and trivial nebentypus.)

Theorem 2. *Assume in addition that*:

- $\bar{\rho}$ *is irreducible*;
- $\bar{\rho}$ *is p-ordinary*, *"nonanomalous" and p-distinguished*:

$$\bar{\rho}|_{D_p} \simeq \begin{pmatrix} \bar{\varepsilon} & * \\ 0 & \bar{\delta} \end{pmatrix},$$

  with $\bar{\varepsilon}, \bar{\delta} : D_p \to \mathbb{F}^\times$ *characters such that* $\bar{\delta}$ *is unramified*, $\bar{\delta}(\mathrm{Frob}_p) \neq \pm 1$ *and* $\bar{\delta} \neq \bar{\varepsilon}$;
- $N(\bar{\rho})^-$ *is the square-free product of an odd number of primes.*

Let $\mathcal{H}^-(\bar{\rho}) := \mathcal{H}^{N(\bar{\rho})^-}(\bar{\rho})$ *consist of all newforms* $f \in \mathcal{H}(\bar{\rho})$ *with* $N_f^- = N(\bar{\rho})^-$, *and fix* $* \in \{\mathrm{alg}, \mathrm{an}\}$. *Then the following hold*:

(1) *For all* $f \in \mathcal{H}^-(\bar{\rho})$, *we have*

$$\mu^*(K_\infty, f) = 0.$$

(2) *Let* $f_1, f_2 \in \mathcal{H}^-(\bar{\rho})$ *lie on the branches* $\mathbb{T}(\mathfrak{a}_1)$, $\mathbb{T}(\mathfrak{a}_2)$ (*defined in* §*1D*), *respectively. Then*

$$\lambda^*(K_\infty, f_1) - \lambda^*(K_\infty, f_2) = \sum_{\ell | N_{f_1}^+ N_{f_2}^+} e_\ell(\mathfrak{a}_2) - e_\ell(\mathfrak{a}_1),$$

where the sum is over the split primes in K *which divide the tame level of* f_1 *or* f_2, *and* $e_\ell(\mathfrak{a}_j)$ *is an explicit nonnegative invariant of the branch* $\mathbb{T}(\mathfrak{a}_j)$ *and the prime* ℓ.

Provided that p splits in K, and under the same hypotheses on $\bar{\rho}$ as in Theorem 2, the work of Skinner and Urban [2014] establishes one of the divisibilities in their "three-variable" Iwasawa main conjecture. Combining their work with our Theorem 2, and making use of the aforementioned results of Bertolini and Darmon [2005] and Pollack and Weston [2011] in weight 2, we obtain many new cases of Conjecture 1 (cf., Corollary 5.5):

Corollary 3. *Suppose that $\bar{\rho}$ is as in Theorem 2 and that p splits in K. Then the anticyclotomic Iwasawa main conjecture holds for every $f \in \mathcal{H}^{-}(\bar{\rho})$ of weight $k \equiv 2 \pmod{p-1}$ and trivial nebentypus.*

Let us briefly explain the new ingredients in the proof of Theorem 2. As it will be clear to the reader, the results contained in Theorem 2 are anticyclotomic analogues of the results of Emerton, Pollack and Weston [Emerton et al. 2006] in the cyclotomic setting. In fact, on the algebraic side the arguments of *loc.cit.* carry over almost verbatim, and our main innovations in this paper are in the development of anticyclotomic analogues of their results on the *analytic* side. Indeed, the analytic results of [Emerton et al. 2006] are based on the study of certain two-variable p-adic L-functions à la Mazur and Kitagawa, whose construction relies on the theory of modular symbols on classical modular curves. In contrast, we need to work on a family of Shimura curves associated with definite quaternion algebras, for which cusps are not available. In the cyclotomic case, modular symbols are useful in two ways: They yield a concrete realization of the degree-one compactly supported cohomology of open modular curves, and provide a powerful tool for studying the arithmetic properties of critical values of the L-functions attached to modular forms. Our basic observation is that in the present anticyclotomic setting, Heegner points on definite Shimura curves provide a similarly convenient way of describing the *central* critical values of the Rankin L-series $L(f/K, \chi, s)$.

Also fundamental for the method of [Emerton et al. 2006] is the possibility to "deform" modular symbols in Hida families. In our anticyclotomic context, the construction of big Heegner points in Hida families was obtained in the work [Longo and Vigni 2011] of one of us in collaboration with Vigni, while the relation between these points and Rankin–Selberg L-values was established in the work [Castella and Longo 2016] by two of us. With these key results at hand, and working over appropriate quotients of the Hecke algebras considered in [Emerton et al. 2006] via the Jacquet–Langlands correspondence, we are then able to develop analogues of the arguments of loc. cit. in our setting, making use of the ramification hypotheses on $\bar{\rho}$ to ensure a multiplicity one property of certain Hecke modules, similarly as in the works of Pollack and Weston [2011] and one of us [Kim 2017].

We conclude this introduction with an overview of the contents of the paper. In the next section, we briefly recall the Hida theory that we need, following the exposition in [Emerton et al. 2006, §1] for the most part. In Section 2, we describe a key extension of the construction of big Heegner points of [Longo and Vigni 2011] to "imprimitive" branches of the Hida family. In Section 3, we construct two-variable p-adic L-functions attached to a Hida family and to each of its irreducible components (or branches), and prove Theorem 3.10 relating the two. This theorem is the key technical result of this paper, and the analytic part of Theorem 2 follows easily from this. In Section 4, we deduce the algebraic part of Theorem 2 using the residual Selmer groups studied in [Pollack and Weston 2011, §3.2]. Finally, in Section 5 we give the applications of our results to the anticyclotomic Iwasawa main conjecture.

1. Hida theory

Throughout this section, we fix a positive integer N admitting a factorization

$$N = N^+N^-$$

with $(N^+, N^-) = 1$ and N^- equal to the square-free product of an *odd* number of primes. We also fix a prime $p \nmid 6N$.

1A. ***Hecke algebras.*** For each integer $k \geqslant 2$, denote by $\mathfrak{h}_{N,r,k}$ the $\mathbb{Z}_p$-algebra generated by the Hecke operators T_ℓ for $\ell \nmid Np$, the operators U_ℓ for $\ell \mid Np$, and the diamond operators $\langle a \rangle$ for $a \in (\mathbb{Z}/p^r\mathbb{Z})^\times$, acting on the space $S_k(\Gamma_{0,1}(N, p^r), \bar{\mathbb{Q}}_p)$ of cusp forms of weight k on $\Gamma_{0,1}(N, p^r) := \Gamma_0(N) \cap \Gamma_1(p^r)$. For $k = 2$, we abbreviate $\mathfrak{h}_{N,r} := \mathfrak{h}_{N,r,2}$.

Let $e^{\mathrm{ord}} := \lim_{n\to\infty} U_p^{n!}$ be Hida's ordinary projector, and define

$$\mathfrak{h}^{\mathrm{ord}}_{N,r,k} := e^{\mathrm{ord}}\mathfrak{h}_{N,r,k}, \qquad \mathfrak{h}^{\mathrm{ord}}_{N,r} := e^{\mathrm{ord}}\mathfrak{h}_{N,r}, \qquad \mathfrak{h}^{\mathrm{ord}}_{N} := \varprojlim_r \mathfrak{h}^{\mathrm{ord}}_{N,r},$$

where the limit is over the projections induced by the natural restriction maps.

Denote by $\mathbb{T}^{N^-}_{N,r,k}$ the quotient of $\mathfrak{h}^{\mathrm{ord}}_{N,r,k}$ acting faithfully on the subspace of $e^{\mathrm{ord}}S_k(\Gamma_{0,1}(N, p^r), \bar{\mathbb{Q}}_p)$ consisting of forms which are new at all primes dividing N^-. Set $\mathbb{T}^{N^-}_{N,r} := \mathbb{T}^{N^-}_{N,r,2}$ and define

$$\mathbb{T}^{N^-}_{N} := \varprojlim_r \mathbb{T}^{N^-}_{N,r}.$$

Each of these Hecke algebras is equipped with natural $\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]$-algebra structures via the diamond operators, and by a well-known result of Hida, $\mathfrak{h}^{\mathrm{ord}}_N$ is finite and flat over $\mathbb{Z}_p[\![1 + p\mathbb{Z}_p]\!]$.

1B. ***Galois representations on Hecke algebras.*** For each positive integer $M \mid N$ we may consider the new quotient $\mathbb{T}^{\mathrm{new}}_M$ of $\mathfrak{h}^{\mathrm{ord}}_M$, and the Galois representation

$$\rho_M : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T}^{\mathrm{new}}_M \otimes \mathcal{L})$$

described in [Emerton et al. 2006, Theorem 2.2.1], where $\mathcal{L}$ denotes the fraction field of $\mathbb{Z}_p[\![1+p\mathbb{Z}_p]\!]$.

Let $\mathbb{T}'_N$ be the $\mathbb{Z}_p[\![1+p\mathbb{Z}_p]\!]$-subalgebra of $\mathbb{T}_N^{N^-}$ generated by the image under the natural projection $\mathfrak{h}_N^{\mathrm{ord}} \to \mathbb{T}_N^{N^-}$ of the Hecke operators of level prime to N. As in [Emerton et al. 2006, Proposition 2.3.2], one can show that the canonical map

$$\mathbb{T}'_N \to \prod_M \mathbb{T}_M^{\mathrm{new}},$$

where the product is over all integers $M \geqslant 1$ with $N^- \mid M \mid N$, becomes an isomorphism after tensoring with $\mathcal{L}$. Taking the product of the Galois representations ρ_M we thus obtain

$$\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T}'_N \otimes \mathcal{L}).$$

For any maximal ideal $\mathfrak{m}$ of $\mathbb{T}'_N$, let $(\mathbb{T}'_N)_{\mathfrak{m}}$ denote the localization of $\mathbb{T}'_N$ at $\mathfrak{m}$ and let

$$\rho_{\mathfrak{m}} : G_{\mathbb{Q}} \to \mathrm{GL}_2((\mathbb{T}'_N)_{\mathfrak{m}} \otimes \mathcal{L})$$

be the resulting Galois representation. If the residual representation $\bar{\rho}_{\mathfrak{m}}$ is irreducible, then $\rho_{\mathfrak{m}}$ admits an integral model (still denoted in the same manner)

$$\rho_{\mathfrak{m}} : G_{\mathbb{Q}} \to \mathrm{GL}_2((\mathbb{T}'_N)_{\mathfrak{m}})$$

which is unique up to isomorphism.

1C. *Residual representations.* Let $\bar{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ be an odd irreducible Galois representation defined over a finite field $\mathbb{F}$ of characteristic $p > 3$. As in the introduction, we assume that $\bar{\rho} \simeq \bar{\rho}_{f_0}$ for some newform f_0 of weight 2, level N, and trivial nebentypus. Consider the following three conditions we may impose on the pair $(\bar{\rho}, N^-)$:

Assumption (SU). (1) $\bar{\rho}$ is *p-ordinary*: the restriction of $\bar{\rho}$ to a decomposition group $D_p \subseteq G_{\mathbb{Q}}$ at p has a one-dimensional unramified quotient over $\mathbb{F}$;

(2) $\bar{\rho}$ is *p-distinguished*: $\bar{\rho}|_{D_p} \sim \left(\begin{smallmatrix} \bar{\varepsilon} & * \\ 0 & \bar{\delta} \end{smallmatrix}\right)$ with $\bar{\varepsilon} \neq \bar{\delta}$;

(3) $\bar{\rho}$ is ramified at every prime $\ell \mid N^-$.

Fix once and for all a representation $\bar{\rho}$ satisfying Assumption (SU), together with a p-stabilization of $\bar{\rho}$ in the sense of [Emerton et al. 2006, Definition 2.2.10]. Let $\overline{V}$ be the two-dimensional $\mathbb{F}$-vector space which affords $\bar{\rho}$, and for any finite set of primes Σ that does not contain p or any factor of N^-, define

$$N(\Sigma) := N(\bar{\rho}) \prod_{\ell \in \Sigma} \ell^{m_\ell}, \tag{1}$$

where $N(\bar{\rho})$ is the tame conductor of $\bar{\rho}$, and $m_\ell := \dim_{\mathbb{F}} \overline{V}_{I_\ell}$.

Remark 1.1. By Assumption (SU) we have the divisibility $N^- \mid N(\bar{\rho})$; we will further assume that $(N^-, N(\bar{\rho})/N^-) = 1$.

Combining [Emerton et al. 2006, Theorem 2.4.1] and [Emerton et al. 2006, Proposition 2.4.2] with the fact that $\bar{\rho}$ is ramified at the primes dividing N^-, one can see that there exist unique maximal ideals $\mathfrak{n}$ and $\mathfrak{m}$ of $\mathbb{T}^{N^-}_{N(\Sigma)}$ and $\mathbb{T}'_{N(\Sigma)}$, respectively, such that

- $\mathfrak{n} \cap \mathbb{T}'_{N(\Sigma)} = \mathfrak{m}$;
- $(\mathbb{T}'_{N(\Sigma)})_{\mathfrak{m}} \simeq (\mathbb{T}^{N^-}_{N(\Sigma)})_{\mathfrak{n}}$ by the natural map on localizations;
- $\bar{\rho}_{\mathfrak{m}} \simeq \bar{\rho}$.

Define the ordinary Hecke algebra $\mathbb{T}_\Sigma$ attached to $\bar{\rho}$ and Σ by

$$\mathbb{T}_\Sigma := (\mathbb{T}'_{N(\Sigma)})_{\mathfrak{m}}.$$

Thus $\mathbb{T}_\Sigma$ is a local factor of $\mathbb{T}'_{N(\Sigma)}$, and we let

$$\rho_\Sigma : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T}_\Sigma)$$

denote the Galois representation deduced from $\rho_{\mathfrak{m}}$.

Adopting the terminology of [Emerton et al. 2006, §2.4], we shall refer to $\mathrm{Spec}(\mathbb{T}_\Sigma)$ as "the Hida family" $\mathcal{H}^-(\bar{\rho})$ attached to $\bar{\rho}$ (and our chosen p-stabilization) that is minimally ramified outside Σ.

Remark 1.2. Note that by Assumption (SU), all the p-stabilized newforms in $\mathcal{H}^-(\bar{\rho})$ have tame level divisible by N^-.

1D. ***Branches of the Hida family.*** If $\mathfrak{a}$ is a minimal prime of $\mathbb{T}_\Sigma$ (for a finite set of primes Σ as above), we put $\mathbb{T}(\mathfrak{a}) := \mathbb{T}_\Sigma/\mathfrak{a}$ and let

$$\rho(\mathfrak{a}) : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{T}(\mathfrak{a}))$$

be the Galois representation induced by ρ_Σ. As in [Emerton et al. 2006, Proposition 2.5.2], one can show that there is a unique divisor $N(\mathfrak{a})$ of $N(\Sigma)$ and a unique minimal prime $\mathfrak{a}' \subseteq \mathbb{T}^{\mathrm{new}}_{N(\mathfrak{a})}$ above $\mathfrak{a}$ such that the diagram

$$\begin{tikzcd}
\mathbb{T}_\Sigma \arrow[r] \arrow[d] & \mathbb{T}'_{N(\Sigma)} \arrow[r] & \prod_{N^- \mid M \mid N(\Sigma)} \mathbb{T}^{\mathrm{new}}_M \arrow[d] \\
\mathbb{T}_\Sigma/\mathfrak{a} \arrow[r, "="] & \mathbb{T}(\mathfrak{a}) \arrow[r] & \mathbb{T}^{\mathrm{new}}_{N(\mathfrak{a})}/\mathfrak{a}'
\end{tikzcd}$$

commutes. We call $N(\mathfrak{a})$ the *tame conductor* of $\mathfrak{a}$ and set

$$\mathbb{T}(\mathfrak{a})^\circ := \mathbb{T}^{\mathrm{new}}_{N(\mathfrak{a})}/\mathfrak{a}'.$$

In particular, note that $N^- \mid N(\mathfrak{a})$ by construction, and that the natural map $\mathbb{T}(\mathfrak{a}) \to \mathbb{T}(\mathfrak{a})^\circ$ is an embedding of local domains.

1E. ***Arithmetic specializations.*** For any finite $\mathbb{Z}_p[[1+p\mathbb{Z}_p]]$-algebra $\mathbb{T}$, we say that a height one prime $\wp$ of $\mathbb{T}$ is an *arithmetic prime* of $\mathbb{T}$ if $\wp$ is the kernel of a $\mathbb{Z}_p$-algebra homomorphism $\mathbb{T} \to \overline{\mathbb{Q}}_p$ such that the composite map

$$1+p\mathbb{Z}_p \to \mathbb{Z}_p[[1+p\mathbb{Z}_p]]^\times \to \mathbb{T}^\times \to \overline{\mathbb{Q}}_p^\times$$

is given by $\gamma \mapsto \gamma^{k-2}$ on some open subgroup of $1+p\mathbb{Z}_p$, for some integer $k \geqslant 2$. We then say that $\wp$ has *weight* k.

Let $\mathfrak{a} \subseteq \mathbb{T}_\Sigma$ be a minimal prime as above. For each $n \geqslant 1$, let $\boldsymbol{a}_n \in \mathbb{T}(\mathfrak{a})^\circ$ be the image of T_n under the natural projection $\mathfrak{h}^{\mathrm{ord}}_{N(\Sigma)} \to \mathbb{T}(\mathfrak{a})^\circ$, and form the q-expansion

$$\boldsymbol{f}(\mathfrak{a}) = \sum_{n\geqslant 1} \boldsymbol{a}_n q^n \in \mathbb{T}(\mathfrak{a})^\circ[[q]].$$

By [Hida 1986, Theorem 1.2], if $\wp$ is an arithmetic prime of $\mathbb{T}(\mathfrak{a})$ of weight k, then there is a unique height one prime $\wp'$ of $\mathbb{T}(\mathfrak{a})^\circ$ such that

$$\boldsymbol{f}_\wp(\mathfrak{a}) := \sum_{n\geqslant 1} (\boldsymbol{a}_n \bmod \wp')q^n \in \mathcal{O}^\circ_\wp[[q]],$$

where $\mathcal{O}^\circ_\wp := \mathbb{T}(\mathfrak{a})^\circ/\wp'$, is the q-expansion of a p-ordinary eigenform $f_\wp$ of weight k and tame level $N(\mathfrak{a})$ (see [Emerton et al. 2006, Proposition 2.5.6]).

2. Big Heegner points

As in Section 1, we fix an integer $N \geqslant 1$ admitting a factorization $N = N^+N^-$ with $(N^+, N^-) = 1$ and N^- equal to the square-free product of an *odd* number of primes, and fix a prime $p \nmid 6N$. Also, we let $K/\mathbb{Q}$ be an imaginary quadratic field of discriminant $-D_K < 0$ prime to Np and such that every prime factor of N^+ (resp. N^-) splits (resp. is inert) in K.

In this section we describe a mild extension of the construction in [Longo and Vigni 2011] (following [Howard 2007]) of big Heegner points attached to K. Indeed, using the results from the preceding section, we can extend the constructions of *loc.cit.* to branches of the Hida family which are *not necessarily* primitive (in the sense of [Hida 1986, §1]). The availability of such an extension is fundamental for the purposes of this paper.

2A. ***Definite Shimura curves.*** Let B be the definite quaternion algebra over $\mathbb{Q}$ of discriminant N^-. We fix once and for all an embedding of $\mathbb{Q}$-algebras $K \hookrightarrow B$, and use it to identity K with a subalgebra of B. Denote by $z \mapsto \bar{z}$ the nontrivial automorphism of K, and choose a basis $\{1, j\}$ of B over K such that

- $j^2 = \beta \in \mathbb{Q}^\times$ with $\beta < 0$;
- $jt = \bar{t}j$ for all $t \in K$;
- $\beta \in (\mathbb{Z}_q^\times)^2$ for $q \mid pN^+$, and $\beta \in \mathbb{Z}_q^\times$ for $q \mid D_K$.

Fix a square-root $\delta_K = \sqrt{-D_K}$, and define $\boldsymbol{\theta} \in K$ by

$$\boldsymbol{\theta} := \tfrac{1}{2}D' + \delta_K, \quad \text{where } D' := \begin{cases} D_K & \text{if } 2 \nmid D_K; \\ \tfrac{1}{2}D_K & \text{if } 2 \mid D_K. \end{cases}$$

Note that $\mathcal{O}_K = \mathbb{Z} + \mathbb{Z}\boldsymbol{\theta}$, and for every prime $q \mid pN^+$, define $i_q : B_q := B \otimes_{\mathbb{Q}} \mathbb{Q}_q \simeq \mathrm{M}_2(\mathbb{Q}_q)$ by

$$i_q(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{Tr}(\boldsymbol{\theta}) & -\mathrm{Nm}(\boldsymbol{\theta}) \\ 1 & 0 \end{pmatrix}, \qquad i_q(j) = \sqrt{\beta}\begin{pmatrix} -1 & \mathrm{Tr}(\boldsymbol{\theta}) \\ 0 & 1 \end{pmatrix},$$

where Tr and Nm are the reduced trace and reduced norm maps on B, respectively. On the other hand, for each prime $q \nmid Np$ we fix any isomorphism $i_q : B_q \simeq \mathrm{M}_2(\mathbb{Q}_q)$ with the property that $i_q(\mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Z}_q) \subset \mathrm{M}_2(\mathbb{Z}_q)$.

For each $r \geqslant 0$, let $R_{N^+,r}$ be the Eichler order of B of level N^+p^r with respect to the above isomorphisms $\{i_q : B_q \simeq \mathrm{M}_2(\mathbb{Q}_q)\}_{q \nmid N^-}$, and let $U_{N^+,r}$ be the compact open subgroup of $\widehat{R}^\times_{N^+,r}$ defined by

$$U_{N^+,r} := \left\{ (x_q)_q \in \widehat{R}^\times_{N^+,r} \,\middle|\, i_p(x_p) \equiv \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix} \pmod{p^r} \right\}.$$

Consider the double coset spaces

$$\widetilde{X}_{N^+,r} = B^\times \setminus (\mathrm{Hom}_{\mathbb{Q}}(K, B) \times \widehat{B}^\times)/U_{N^+,r}, \tag{2}$$

where $b \in B^\times$ acts on $(\Psi, g) \in \mathrm{Hom}_{\mathbb{Q}}(K, B) \times \widehat{B}^\times$ by

$$b \cdot (\Psi, g) = (b\Psi b^{-1}, bg)$$

and $U_{N^+,r}$ acts on $\widehat{B}^\times$ by right multiplication. As is well known (see, e.g., [Longo and Vigni 2011, §2.1]), $\widetilde{X}_{N^+,r}$ may be naturally identified with the set of K-rational points of certain genus zero curves defined over $\mathbb{Q}$. Nonetheless, there is a nontrivial Galois action on $\widetilde{X}_{N^+,r}$ defined as follows: If $\sigma \in \mathrm{Gal}(K^{\mathrm{ab}}/K)$ and $P \in \widetilde{X}_{N^+,r}$ is the class of a pair (Ψ, g), then

$$P^\sigma := [(\Psi, \widehat{\Psi}(a)g)],$$

where $a \in K^\times \setminus \widehat{K}^\times$ is chosen so that $\mathrm{rec}_K(a) = \sigma$. It will be convenient to extend this action to an action of $G_K := \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ in the obvious manner.

Finally, we note that $\widetilde{X}_{N^+,r}$ is also equipped with standard actions of U_p, Hecke operators T_ℓ for $\ell \nmid Np$, and diamond operators $\langle d \rangle$ for $d \in (\mathbb{Z}/p^r\mathbb{Z})^\times$ (see [Longo and Vigni 2011, §2.4], for example).

2B. ***Compatible systems of Heegner points.*** For each integer $c \geqslant 1$, let $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$ be the order of K of conductor c.

Definition 2.1. We say that a point $P \in \widetilde{X}_{N^+,r}$ is a *Heegner point of conductor c* if P is the class of a pair (Ψ, g) with

$$\Psi(\mathcal{O}_c) = \Psi(K) \cap (B \cap g\widehat{R}_{N^+,r}g^{-1})$$

and

$$\Psi_p((\mathcal{O}_c \otimes \mathbb{Z}_p)^\times \cap (1 + p^r\mathcal{O}_K \otimes \mathbb{Z}_p)^\times) = \Psi_p((\mathcal{O}_c \otimes \mathbb{Z}_p)^\times) \cap g_p U_{N^+,r,p} g_p^{-1},$$

where $U_{N^+,r,p}$ denotes the p-component of $U_{N^+,r}$.

Fix a decomposition $N^+\mathcal{O}_K = \mathfrak{N}^+\overline{\mathfrak{N}^+}$, and for each prime $q \neq p$ define

- $\varsigma_q = 1$, if $q \nmid N^+$;
- $\varsigma_q = \delta_K^{-1}\begin{pmatrix}\boldsymbol{\theta} & \bar{\boldsymbol{\theta}}\\ 1 & 1\end{pmatrix} \in \mathrm{GL}_2(K_\mathfrak{q}) = \mathrm{GL}_2(\mathbb{Q}_q)$, if $q = \mathfrak{q}\bar{\mathfrak{q}}$ splits with $\mathfrak{q} \mid \mathfrak{N}^+$,

and for each $s \geqslant 0$, let

- $\varsigma_p^{(s)} = \begin{pmatrix}\boldsymbol{\theta} & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}p^s & 0\\ 0 & 1\end{pmatrix} \in \mathrm{GL}_2(K_\mathfrak{p}) = \mathrm{GL}_2(\mathbb{Q}_p)$, if $p = \mathfrak{p}\bar{\mathfrak{p}}$ splits in K;
- $\varsigma_p^{(s)} = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}p^s & 0\\ 0 & 1\end{pmatrix}$, if p is inert in K.

Set $\varsigma^{(s)} := \varsigma_p^{(s)}\prod_{q\neq p}\varsigma_q$, viewed as an element in $\widehat{B}^\times$ via the isomorphisms $\{i_q : B_q \simeq M_2(\mathbb{Q}_q)\}_{q\nmid N^-}$ introduced in Section 2A. Let $\imath_K : K \hookrightarrow B$ be the inclusion. Then one easily checks (see [Castella and Longo 2016, Theorem 1.2]) that for all $n, r \geqslant 0$ the points

$$\widetilde{P}_{p^n,r} := [(\imath_K, \varsigma^{(n+r)})] \in \widetilde{X}_{N^+,r}$$

are Heegner points of conductor p^{n+r} with the following properties:

- *Field of definition*: $\widetilde{P}_{p^n,r} \in H^0(L_{p^n,r}, \widetilde{X}_{N^+,r})$, where $L_{p^n,r} := H_{p^{n+r}}(\boldsymbol{\mu}_{p^r})$ and H_c is the ring class field of K of conductor c.
- *Galois equivariance*: for all $\sigma \in \mathrm{Gal}(L_{p^n,r}/H_{p^{n+r}})$, we have

  $$\widetilde{P}^\sigma_{p^n,r} = \langle\vartheta(\sigma)\rangle \cdot \widetilde{P}_{p^n,r},$$

  where $\vartheta : \mathrm{Gal}(L_{p^n,r}/H_{p^{n+r}}) \to \mathbb{Z}_p^\times/\{\pm 1\}$ is such that $\vartheta^2 = \varepsilon_{\mathrm{cyc}}$.
- *Horizontal compatibility*: if $r > 1$, then

  $$\sum_{\sigma\in\mathrm{Gal}(L_{p^n,r}/L_{p^{n-1},r})} \widetilde{\alpha}_r(\widetilde{P}^\sigma_{p^n,r}) = U_p \cdot \widetilde{P}_{p^n,r-1},$$

  where $\widetilde{\alpha}_r : \widetilde{X}_{N^+,r} \to \widetilde{X}_{N^+,r-1}$ is the map induced by the inclusion $U_{N^+,r} \subseteq U_{N^+,r-1}$.

- *Vertical Compatibility*: if $n > 0$, then

$$\sum_{\sigma \in \mathrm{Gal}(L_{p^n,r}/L_{p^{n-1},r})} \widetilde{P}^{\sigma}_{p^n,r} = U_p \cdot \widetilde{P}_{p^{n-1},r}.$$

Remark 2.2. We will only consider the points $\widetilde{P}_{p^n,r}$ for a fixed a value of N^- (which amounts to fixing the quaternion algebra $B/\mathbb{Q}$), but it will be fundamental to consider *different* values of N^+, and the relations between the corresponding $\widetilde{P}_{p^n,r}$ (which clearly depend on N^+) under various degeneracy maps.

2C. *Critical character.* Factor the p-adic cyclotomic character as

$$\varepsilon_{\mathrm{cyc}} = \varepsilon_{\mathrm{tame}} \cdot \varepsilon_{\mathrm{wild}} : G_{\mathbb{Q}} \to \mathbb{Z}_p^{\times} \simeq \boldsymbol{\mu}_{p-1} \times (1 + p\mathbb{Z}_p)$$

and define the *critical character* $\Theta : G_{\mathbb{Q}} \to \mathbb{Z}_p[\![1 + p\mathbb{Z}_p]\!]^{\times}$ by

$$\Theta(\sigma) = [\varepsilon^{1/2}_{\mathrm{wild}}(\sigma)], \tag{3}$$

where $\varepsilon^{1/2}_{\mathrm{wild}}$ is the unique square root of $\varepsilon_{\mathrm{wild}}$ taking values in $1 + p\mathbb{Z}_p$, and the map $[\,\cdot\,] : 1 + p\mathbb{Z}_p \to \mathbb{Z}_p[\![1 + p\mathbb{Z}_p]\!]^{\times}$ is given by the inclusion as group-like elements.

2D. *Big Heegner points.* Recall the Shimura curves $\widetilde{X}_{N^+,p^r}$ from Section 2A, and set

$$\mathfrak{D}_{N^+,r} := e^{\mathrm{ord}}(\mathrm{Div}(\widetilde{X}_{N^+,r}) \otimes_{\mathbb{Z}} \mathbb{Z}_p).$$

By the Jacquet–Langlands correspondence, $\mathfrak{D}_{N^+,r}$ is naturally endowed with an action of the Hecke algebra $\mathbb{T}^{N^-}_{N,r}$. Let $(\mathbb{T}^{N^-}_{N,r})^{\dagger}$ be the free $\mathbb{T}^{N^-}_{N,r}$-module of rank one equipped with the Galois action via the inverse of the critical character Θ, and set

$$\mathfrak{D}^{\dagger}_{N^+,r} := \mathfrak{D}_{N^+,r} \otimes_{\mathbb{T}^{N^-}_{N,r}} (\mathbb{T}^{N^-}_{N,r})^{\dagger}.$$

Let $\widetilde{P}_{p^n,r} \in \widetilde{X}_{N^+,r}$ be the system of Heegner points of Section 2B, and denote by $\mathcal{P}_{p^n,r}$ the image of $e^{\mathrm{ord}}\widetilde{P}_{p^n,r}$ in $\mathfrak{D}_{N^+,r}$. By the Galois equivariance of $\widetilde{P}_{p^n,r}$ (see [Longo and Vigni 2011, §7.1]), we have

$$\mathcal{P}^{\sigma}_{p^n,r} = \Theta(\sigma) \cdot \mathcal{P}_{p^n,r}$$

for all $\sigma \in \mathrm{Gal}(L_{p^n,r}/H_{p^{n+r}})$, and hence $\mathcal{P}_{p^n,r}$ defines an element

$$\mathcal{P}_{p^n,r} \otimes \zeta_r \in H^0(H_{p^{n+r}}, \mathfrak{D}^{\dagger}_{N^+,r}). \tag{4}$$

In the next section we shall see how this system of points, for varying n and r, can be used to construct various anticyclotomic p-adic L-functions.

3. Anticyclotomic p-adic L-functions

3A. *Multiplicity one.* Keep the notation introduced in Section 2. For each integer $k \geqslant 2$, denote by $L_k(R)$ the set of polynomials of degree less than or equal to $k-2$ with coefficients in a ring R, and define

$$\mathfrak{J}_{N^+,r,k} := e^{\mathrm{ord}} H_0(\widetilde{X}_{N^+,r}, \mathcal{L}_k(\mathbb{Z}_p)),$$

where $\mathcal{L}_k(\mathbb{Z}_p)$ is the local system on $\widetilde{X}_{N^+,r}$ associated with $L_k(\mathbb{Z}_p)$. The module $\mathfrak{J}_{N^+,r,k}$ is endowed with an action of the Hecke algebra $\mathbb{T}^{N^-}_{N,r,k}$ and with perfect "intersection pairing":

$$\langle\,\cdot\,,\,\cdot\,\rangle_k : \mathfrak{J}_{N^+,r,k} \times \mathfrak{J}_{N^+,r,k} \to \mathbb{Q}_p \tag{5}$$

(see [Chida and Hsieh 2016, Equation (3.9)]) with respect to which the Hecke operators are self-adjoint.

Theorem 3.1. *Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}^{N^-}_{N,r,k}$ whose residual representation is irreducible and satisfies Assumption (SU). Then $(\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}}$ is free of rank one over $(\mathbb{T}^{N^-}_{N,r,k})_{\mathfrak{m}}$. In particular, there is a $(\mathbb{T}^{N^-}_{N,r,k})_{\mathfrak{m}}$-module isomorphism*

$$(\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}} \overset{\alpha_{N,r,k}}{\simeq} (\mathbb{T}^{N^-}_{N,r,k})_{\mathfrak{m}}.$$

Proof. If $k=2$ and $r=1$, this follows by combining Theorem 6.2 and Proposition 6.5 of [Pollack and Weston 2011]. The general case will be deduced from this case in Section 3C using Hida theory. □

Let $f \in S_k(\Gamma_{0,1}(N, p^r))$ be an N^--new eigenform, and suppose that $\mathfrak{m}$ is the maximal ideal of $\mathbb{T}^{N^-}_{N,r,k}$ containing the kernel of the associated $\mathbb{Z}_p$-algebra homomorphism

$$\pi_f : (\mathbb{T}^{N^-}_{N,r,k})_{\mathfrak{m}} \to \mathcal{O},$$

where $\mathcal{O}$ is the finite extension of $\mathbb{Z}_p$ generated by the Fourier coefficients of f. Composing π_f with an isomorphism $\alpha_{N,r,k}$ as in Theorem 3.1, we obtain an $\mathcal{O}$-valued functional

$$\psi_f : (\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}} \to \mathcal{O}.$$

By the duality (5), the map ψ_f corresponds to a generator g_f of the π_f-isotypical component of $\mathfrak{J}_{N^+,r,k}$, and following [Pollack and Weston 2011, §2.1] and [Chida and Hsieh 2016, §4.1] we define the *Gross period* Ω_{f,N^-} attached to f by

$$\Omega_{f,N^-} := \frac{(f,f)_{\Gamma_0(N)}}{\langle g_f, g_f\rangle_k}. \tag{6}$$

Remark 3.2. By Vatsal's work [2003] (see also [Pollack and Weston 2011, Theorem 2.3] and [Chida and Hsieh 2016, §5.4]), the anticyclotomic p-adic L-functions

$L_p(f/K)$ in Theorem 3.14 below (normalized by the complex period Ω_{f,N^-}) have vanishing μ-invariant. The preceding uniform description of ψ_f for all f with a common maximal ideal $\mathfrak{m}$ will allow us to show that this property is preserved in Hida families.

3B. ***One-variable p-adic L-functions.*** Denote by Γ the Galois group of the anticyclotomic $\mathbb{Z}_p$-extension K_∞/K. For each n, let $K_n \subset K_\infty$ be defined by $\mathrm{Gal}(K_n/K) \simeq \mathbb{Z}/p^n\mathbb{Z}$ and let Γ_n be the subgroup of Γ such that $\Gamma/\Gamma_n \simeq \mathrm{Gal}(K_n/K)$.

Let $\mathcal{P}_{p^{n+1},r} \otimes \zeta_r \in H^0(H_{p^{n+1+r}}, \mathfrak{D}^\dagger_{N^+,r})$ be the Heegner point of conductor p^{n+1}, and define

$$\mathcal{Q}_{n,r} := \mathrm{Cor}_{H_{p^{n+1+r}}/K_n}(\mathcal{P}_{p^{n+1},r} \otimes \zeta_r) \in H^0(K_n, \mathfrak{D}^\dagger_{N^+,r}); \tag{7}$$

with a slight abuse of notation, we also denote by $\mathcal{Q}_{n,r}$ its image under the natural map

$$H^0(K_n, \mathfrak{D}^\dagger_{N^+,r}) \xrightarrow{\subseteq} \mathfrak{D}_{N^+,r} \longrightarrow \mathfrak{J}_{N^+,r}$$

composed with localization at $\mathfrak{m}$, where $\mathfrak{J}_{N^+,r} := \mathfrak{J}_{N^+,r,2}$.

Definition 3.3. For any open subset $\sigma\Gamma_n$ of Γ, define

$$\mu_r(\sigma\Gamma_n) := U_p^{-n} \cdot \mathcal{Q}^\sigma_{n,r} \in (\mathfrak{J}_{N^+,r})_{\mathfrak{m}}.$$

Proposition 3.4. *The rule μ_r is a measure on Γ.*

Proof. This follows immediately from the "horizontal compatibility" of Heegner points. □

3C. ***Gross periods in Hida families.*** Keep the notation of Section 3A, and let

$$(\mathfrak{J}_{N^+})_{\mathfrak{m}} := \varprojlim_r (\mathfrak{J}_{N^+,r})_{\mathfrak{m}},$$

which is naturally equipped with an action of the big Hecke algebra $\mathbb{T}_N^{N^-} = \varprojlim_r \mathbb{T}^{N^-}_{N,r}$.

Theorem 3.5. *Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}_N^{N^-}$ whose residual representation is irreducible and satisfies Assumption (SU). Then $(\mathfrak{J}_{N^+})_{\mathfrak{m}}$ is free of rank one over $(\mathbb{T}_N^{N^-})_{\mathfrak{m}}$. In particular, there is a $(\mathbb{T}_N^{N^-})_{\mathfrak{m}}$-module isomorphism*

$$(\mathfrak{J}_{N^+})_{\mathfrak{m}} \overset{\alpha_N}{\simeq} (\mathbb{T}_N^{N^-})_{\mathfrak{m}}.$$

Proof. As in [Emerton et al. 2006, Proposition 3.3.1]. Note that the version of Hida's control theorem in our context is provided by [Hida 1988, Theorem 9.4]. □

We can now conclude the proof of Theorem 3.1 just as in [Emerton et al. 2006, §3.3]. For the convenience of the reader, we include the argument here.

Proof of Theorem 3.1. Let $\wp_{N,r,k}$ be the product of all the arithmetic primes of $\mathbb{T}_N^{N^-}$ of weight k which become trivial upon restriction to $1+p^r\mathbb{Z}_p$. By [Hida 1988, Theorem 9.4], we then have

$$(\mathfrak{J}_{N^+})_{\mathfrak{m}} \otimes \mathbb{T}_N^{N^-}/\wp_{N,r,k} \simeq (\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}_{r,k}}, \tag{8}$$

where $\mathfrak{m}_{r,k}$ is the maximal ideal of $\mathbb{T}_{N,r,k}^{N^-}$ induced by $\mathfrak{m}$. Since $(\mathfrak{J}_{N^+})_{\mathfrak{m}}$ is free of rank one over $\mathbb{T}_N^{N^-}$ by Theorem 3.5, it follows that $(\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}_{r,k}}$ is free of rank one over $\mathbb{T}_N^{N^-}/\wp_{N,r,k} \simeq \mathbb{T}_{N,r,k}^{N^-}$, as was to be shown. □

Remark 3.6. In the above proofs we made crucial use of [Hida 1988, Theorem 9.4], which is stated in the context of totally definite quaternion algebras which are unramified at all finite places, since this is the only relevant case for the study of Hilbert modular forms over totally real number fields of even degree. However, the proofs immediately extend to the (simpler) situation of definite quaternion algebras over $\mathbb{Q}$.

3D. ***Two-variable p-adic L-functions.*** By the "vertical compatibility" satisfied by Heegner points, the points

$$U_p^{-r} \cdot \mathcal{Q}_{n,r} \in (\mathfrak{J}_{N^+,r})_{\mathfrak{m}}$$

are compatible for varying r, thus defining an element

$$\mathcal{Q}_n := \varprojlim_r U_p^{-r} \cdot \mathcal{Q}_{n,r} \in (\mathfrak{J}_{N^+})_{\mathfrak{m}}.$$

Definition 3.7. For any open subset $\sigma\Gamma_n$ of Γ, define

$$\mu(\sigma\Gamma_n) := U_p^{-n} \cdot \mathcal{Q}_n^{\sigma} \in (\mathfrak{J}_{N^+})_{\mathfrak{m}}.$$

Proposition 3.8. *The rule μ is a measure on Γ.*

Proof. This follows immediately from the "horizontal compatibility" of Heegner points. □

Upon the choice of an isomorphism α_N as in Theorem 3.5, we may regard μ as an element

$$\mathcal{L}(\mathfrak{m}, N) \in (\mathbb{T}_N^{N^-})_{\mathfrak{m}} \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[\![\Gamma]\!].$$

Denoting by $\mathcal{L}(\mathfrak{m}, N)^*$ the image of $\mathcal{L}(\mathfrak{m}, N)$ under the involution induced by $\gamma \mapsto \gamma^{-1}$ on group-like elements, we set

$$L(\mathfrak{m}, N) := \mathcal{L}(\mathfrak{m}, N) \cdot \mathcal{L}(\mathfrak{m}, N)^*,$$

to which we will refer as the *two-variable p-adic L-function attached to* $(\mathbb{T}_N^{N^-})_{\mathfrak{m}}$.

3E. ***Two-variable p-adic L-functions on branches of the Hida family.*** Let $\mathbb{T}_\Sigma$ be the universal p-ordinary Hecke algebra

$$\mathbb{T}_\Sigma := (\mathbb{T}'_{N(\Sigma)})_{\mathfrak{m}} \simeq (\mathbb{T}^{N^-}_{N(\Sigma)})_{\mathfrak{n}} \tag{9}$$

associated with a mod p representation $\bar{\rho}$ and a finite set of primes Σ as in Section 1C.

Remark 3.9. Recall that $N^- \mid N(\bar{\rho})$ by Assumption (SU). Throughout the following, it will be further assumed that every prime factor of $N(\Sigma)/N^-$ splits in K. In particular, every prime $\ell \in \Sigma$ splits in K, and any $f \in \mathcal{H}^-(\bar{\rho}) = \mathrm{Spec}(\mathbb{T}_\Sigma)$ has tame level N_f with

$$N_f^- = N(\bar{\rho})^- = N^-.$$

The construction of the preceding section produces a two-variable p-adic L-function

$$L(\mathfrak{n}, N(\Sigma)) \in (\mathbb{T}^{N^-}_{N(\Sigma)})_{\mathfrak{n}} \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[[\Gamma]],$$

which combined with the isomorphism (9) yields an element

$$L_\Sigma(\bar{\rho}) \in \mathbb{T}_\Sigma \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[[\Gamma]].$$

If $\mathfrak{a}$ is a minimal prime of $\mathbb{T}_\Sigma$, we thus obtain an element

$$L_\Sigma(\bar{\rho}, \mathfrak{a}) \in \mathbb{T}(\mathfrak{a})^\circ \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[[\Gamma]]$$

by reducing $L_\Sigma(\bar{\rho})$ mod $\mathfrak{a}$ (see Section 1D). On the other hand, if we let $\mathfrak{m}$ denote the inverse image of the maximal ideal of $\mathbb{T}(\mathfrak{a})^\circ$ under the composite surjection

$$\mathbb{T}^{N^-}_{N(\mathfrak{a})} \to \mathbb{T}^{\mathrm{new}}_{N(\mathfrak{a})} \to \mathbb{T}^{\mathrm{new}}_{N(\mathfrak{a})}/\mathfrak{a}' = \mathbb{T}(\mathfrak{a})^\circ, \tag{10}$$

the construction of the preceding section yields an L-function

$$L(\mathfrak{m}, N(\mathfrak{a})) \in (\mathbb{T}^{N^-}_{N(\mathfrak{a})})_{\mathfrak{m}} \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[[\Gamma]]$$

giving rise, via (10), to a second element

$$L(\bar{\rho}, \mathfrak{a}) \in \mathbb{T}(\mathfrak{a})^\circ \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[[\Gamma]].$$

It is natural to compare $L_\Sigma(\bar{\rho}, \mathfrak{a})$ and $L(\bar{\rho}, \mathfrak{a})$, a task that is carried out in the next section, and provides the key for understanding the variation of *analytic* Iwasawa invariants.

3F. *Comparison.* Write $\Sigma = \{\ell_1, \dots, \ell_n\}$ and for each $\ell = \ell_i \in \Sigma$, let e_ℓ be the valuation of $N(\Sigma)/N(\mathfrak{a})$ at ℓ, and define the reciprocal Euler factor $E_\ell(\mathfrak{a}, X) \in \mathbb{T}(\mathfrak{a})^\circ[X]$ by

$$E_\ell(\mathfrak{a}, X) := \begin{cases} 1 & \text{if } e_\ell = 0; \\ 1 - (T_\ell \bmod \mathfrak{a}')\Theta^{-1}(\ell)X & \text{if } e_\ell = 1; \\ 1 - (T_\ell \bmod \mathfrak{a}')\Theta^{-1}(\ell)X + \ell X^2 & \text{if } e_\ell = 2. \end{cases}$$

Also, writing $\ell = \mathfrak{l}\bar{\mathfrak{l}}$, define $E_\ell(\mathfrak{a}) \in \mathbb{T}(\mathfrak{a})^\circ \hat{\otimes}_{\mathbb{Z}_p} \mathbb{Z}_p[\![\Gamma]\!]$ by

$$E_\ell(\mathfrak{a}) := E_\ell(\mathfrak{a}, \ell^{-1}\gamma_{\mathfrak{l}}) \cdot E_\ell(\mathfrak{a}, \ell^{-1}\gamma_{\bar{\mathfrak{l}}}), \tag{11}$$

where $\gamma_{\mathfrak{l}}, \gamma_{\bar{\mathfrak{l}}}$ are arithmetic Frobenius maps at $\mathfrak{l}, \bar{\mathfrak{l}}$ in Γ, respectively, and put $E_\Sigma(\mathfrak{a}) := \prod_{\ell \in \Sigma} E_\ell(\mathfrak{a})$.

Recall that $N^- \mid N(\mathfrak{a}) \mid N(\Sigma)$ and set

$$N(\mathfrak{a})^+ := N(\mathfrak{a})/N^-, \qquad N(\Sigma)^+ := N(\Sigma)/N^-,$$

both of which consist entirely of prime factors which split in K. The purpose of this section is to prove the following result.

Theorem 3.10. *There is an isomorphism of $\mathbb{T}(\mathfrak{a})^\circ$-modules*

$$\mathbb{T}(\mathfrak{a})^\circ \otimes_{(\mathbb{T}^{N^-}_{N(\Sigma)})_{\mathfrak{n}}} (\mathfrak{J}_{N(\Sigma)^+})_{\mathfrak{n}} \simeq \mathbb{T}(\mathfrak{a})^\circ \otimes_{(\mathbb{T}^{N^-}_{N(\mathfrak{a})})_{\mathfrak{m}}} (\mathfrak{J}_{N(\mathfrak{a})^+})_{\mathfrak{m}}$$

such that the map induced on the corresponding spaces of measures valued in these modules sends $L_\Sigma(\bar{\rho}, \mathfrak{a})$ to $E_\Sigma(\mathfrak{a}) \cdot L(\bar{\rho}, \mathfrak{a})$.

Proof. The proof follows closely the constructions and arguments in [Emerton et al. 2006, §3.8], suitably adapted to the quaternionic setting at hand. Let $r \geqslant 1$. If M is any positive integer with $(M, pN^-) = 1$, and $d' \mid d$ are divisors of M, we have degeneracy maps

$$B_{d,d'} : \widetilde{X}_{M,r} \to \widetilde{X}_{M/d,r}$$

induced by $(\Psi, g) \mapsto (\Psi, \pi_{d'} g)$, where $\pi_{d'} \in \widehat{B}^\times$ has local component $\left(\begin{smallmatrix} 1 & 0 \\ 0 & \ell^{\mathrm{val}_\ell(d')} \end{smallmatrix}\right)$ at every prime $\ell \mid d'$ and 1 outside d'. We thus obtain a map on homology

$$(B_{d,d'})_* : e^{\mathrm{ord}} H_0(\widetilde{X}_{M,r}, \mathbb{Z}_p) \to e^{\mathrm{ord}} H_0(\widetilde{X}_{M/d,r}, \mathbb{Z}_p)$$

and we may define

$$\epsilon_r : e^{\mathrm{ord}} H_0(\widetilde{X}_{N(\Sigma)^+,r}, \mathbb{Z}_p) \to e^{\mathrm{ord}} H_0(\widetilde{X}_{N(\mathfrak{a})^+,r}, \mathbb{Z}_p) \tag{12}$$

by $\epsilon_r := \epsilon(\ell_n) \circ \cdots \circ \epsilon(\ell_1)$, where for every $\ell = \ell_i \in \Sigma$ we put

$$\epsilon(\ell) := \begin{cases} 1 & \text{if } e_\ell = 0; \\ (B_{\ell,1})_* - (B_{\ell,\ell})_* \ell^{-1} T_\ell & \text{if } e_\ell = 1; \\ (B_{\ell^2,1})_* - (B_{\ell^2,\ell})_* \ell^{-1} T_\ell + (B_{\ell^2,\ell^2})_* \ell^{-1} \langle \ell \rangle_{N(\mathfrak{a})p} & \text{if } e_\ell = 2. \end{cases}$$

As before, let M be a positive integer with $(M, pN^-) = 1$ all of whose prime factors split in K, and let $\ell \nmid Mp$ be a prime which also splits in K. We shall adopt the following simplifying notation for the system of points $\widetilde{P}_{p^n,r} \in \widetilde{X}_{N^+,r}$ constructed in Section 2B:

$$P := \widetilde{P}^{(M)}_{p^n,r} \in \widetilde{X}_{M,r}, \quad P^{(\ell)} := \widetilde{P}^{(M\ell)}_{p^n,r} \in \widetilde{X}_{M\ell,r}, \quad P^{(\ell^2)} := \widetilde{P}^{(M\ell^2)}_{p^n,r} \in \widetilde{X}_{M\ell^2,r}.$$

It is easy to check that we have the following relations in $\widetilde{X}_{M,r}$:

$$(B_{\ell,1})_*(P^{(\ell)}) = P, \qquad (B_{\ell,\ell})_*(P^{(\ell)}) = P^{\sigma_{\mathfrak{l}}}, \quad (B_{\ell^2,1})_*(P^{(\ell^2)}) = P,$$
$$(B_{\ell^2,\ell})_*(P^{(\ell^2)}) = P^{\sigma_{\mathfrak{l}}}, \quad (B_{\ell^2,\ell^2})_*(P^{(\ell^2)}) = P^{\sigma_{\mathfrak{l}}^2},$$

where $\sigma_{\mathfrak{l}} \in \mathrm{Gal}(L_{p^n,r}/K)$ is a Frobenius element at a prime $\mathfrak{l} \mid \ell$. Letting $\mathcal{P}$ denote the image of $e^{\mathrm{ord}}P$ in $\mathfrak{D}_{M,r}$, and defining $\mathcal{P}^{(\ell)} \in \mathfrak{D}_{M\ell,r}$ and $\mathcal{P}^{(\ell^2)} \in \mathfrak{D}_{M\ell^2,r}$ similarly, it follows that

$$\begin{aligned}
(B_{\ell,1})_*(\mathcal{P}^{(\ell)} \otimes \zeta_r) &= \mathcal{P} \otimes \zeta_r,\\
(B_{\ell,\ell})_*(\mathcal{P}^{(\ell)} \otimes \zeta_r) &= \mathcal{P}^{\sigma_{\mathfrak{l}}} \otimes \zeta_r \;= \Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot (\mathcal{P} \otimes \zeta_r)^{\sigma_{\mathfrak{l}}},\\
(B_{\ell^2,1})_*(\mathcal{P}^{(\ell^2)} \otimes \zeta_r) &= \mathcal{P} \otimes \zeta_r,\\
(B_{\ell^2,\ell})_*(\mathcal{P}^{(\ell^2)} \otimes \zeta_r) &= \mathcal{P}^{\sigma_{\mathfrak{l}}} \otimes \zeta_r \;= \Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot (\mathcal{P} \otimes \zeta_r)^{\sigma_{\mathfrak{l}}},\\
(B_{\ell^2,\ell^2})_*(\mathcal{P}^{(\ell^2)} \otimes \zeta_r) &= \mathcal{P}^{\sigma_{\mathfrak{l}}^2} \otimes \zeta_r = \Theta^{-2}(\sigma_{\mathfrak{l}}) \cdot (\mathcal{P} \otimes \zeta_r)^{\sigma_{\mathfrak{l}}}
\end{aligned}$$

as elements in $\mathfrak{D}^{\dagger}_{M,r}$. Finally, setting $\mathcal{Q} := \mathrm{Cor}_{H_{p^{n+1+r}}/K_n}(\mathcal{P}) \in H^0(K_n, \mathfrak{D}^{\dagger}_{M,r})$, and defining $\mathcal{Q}^{(\ell)} \in H^0(K_n, \mathfrak{D}^{\dagger}_{M\ell,r})$ and $\mathcal{Q}^{(\ell^2)} \in H^0(K_n, \mathfrak{D}^{\dagger}_{M\ell^2,r})$ similarly, we see that

$$(B_{\ell,1})_*(\mathcal{Q}^{(\ell)}) = \mathcal{Q}, \qquad (B_{\ell,\ell})_*(\mathcal{Q}^{(\ell)}) = \Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot \mathcal{Q}^{\sigma_{\mathfrak{l}}},$$
$$(B_{\ell^2,1})_*(\mathcal{Q}^{(\ell^2)}) = \mathcal{Q},$$
$$(B_{\ell^2,\ell})_*(\mathcal{Q}^{(\ell^2)}) = \Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot \mathcal{Q}^{\sigma_{\mathfrak{l}}}, \quad (B_{\ell^2,\ell^2})_*(\mathcal{Q}^{(\ell^2)}) = \Theta^{-2}(\sigma_{\mathfrak{l}}) \cdot \mathcal{Q}^{\sigma_{\mathfrak{l}}^2}$$

in $H^0(K_n, \mathfrak{D}^{\dagger}_{M,r})$. Each of these equalities is checked by an explicit calculation. For example, for the second one:

$$\begin{aligned}
(B_{\ell,\ell})_*(\mathcal{Q}^{(\ell)}) &= (B_{\ell,\ell})_*(\mathrm{Cor}_{H_{p^{n+1+r}}/K_n}(\mathcal{P}^{(\ell)} \otimes \zeta_r))\\
&= (B_{\ell,\ell})_*\Bigg(\Bigg(\sum_{\sigma \in \mathrm{Gal}(H_{p^{n+1+r}}/K_n)} \Theta(\tilde{\sigma}^{-1}) \cdot (\mathcal{P}^{(\ell)})^{\tilde{\sigma}}\Bigg) \otimes \zeta_r\Bigg)\\
&= \sum_{\sigma \in \mathrm{Gal}(H_{p^{n+1+r}}/K_n)} \Theta(\tilde{\sigma}^{-1}) \cdot (B_{\ell,\ell})_*((\mathcal{P}^{(\ell)})^{\tilde{\sigma}} \otimes \zeta_r)\\
&= \sum_{\sigma \in \mathrm{Gal}(H_{p^{n+1+r}}/K_n)} \Theta(\tilde{\sigma}^{-1})\Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot (\mathcal{P}^{\tilde{\sigma}} \otimes \zeta_r)^{\sigma_{\mathfrak{l}}}\\
&= \Theta^{-1}(\sigma_{\mathfrak{l}}) \cdot \mathcal{Q}^{\sigma_{\mathfrak{l}}}.
\end{aligned}$$

Now let $\mathcal{Q}_{n,r} \in \mathfrak{J}_{N(\Sigma)^+,r}$ be as in (7) with $N = N(\Sigma)$. Using the above formulae, we easily see that of any finite order character χ of Γ of conductor p^n, the effect of ϵ_r on the element $\sum_{\sigma\in\Gamma/\Gamma_n} \chi(\sigma)\mathcal{Q}_{n,r}^{\sigma}$ is given by multiplication by

$$\prod_{e_{\ell_i}=1}(1-(\chi\Theta)^{-1}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}T_{\ell_i})\prod_{e_{\ell_i}=2}(1-(\chi\Theta)^{-1}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}T_{\ell_i}+(\chi\Theta)^{-2}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}\langle\ell_i\rangle_{N(\mathfrak{a})p}).$$

Similarly, we see ϵ_r has the effect of multiplying the element $\sum\limits_{\sigma\in\Gamma/\Gamma_n} \chi^{-1}(\sigma)\mathcal{Q}_{n,r}^{\sigma}$ by

$$\prod_{e_{\ell_i}=1}(1-(\chi^{-1}\Theta)^{-1}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}T_{\ell_i})\prod_{e_{\ell_i}=2}(1-(\chi^{-1}\Theta)^{-1}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}T_{\ell_i}+(\chi^{-1}\Theta)^{-2}(\sigma_{\mathfrak{l}_i})\ell_i^{-1}\langle\ell_i\rangle_{N(\mathfrak{a})p}).$$

Hence, using the relations

$$\chi(\sigma_{\bar{\mathfrak{l}}_i}) = \chi^{-1}(\sigma_{\mathfrak{l}_i}), \qquad \Theta(\sigma_{\mathfrak{l}_i}) = \Theta(\sigma_{\bar{\mathfrak{l}}_i}) = \theta(\ell_i), \qquad \theta^2(\ell_i) = \langle\ell_i\rangle_{N(\mathfrak{a})p},$$

it follows that the effect of ϵ_r on the product of the above two elements is given by multiplication by

$$\prod_{\substack{\mathfrak{l}_i|\ell_i\\ e_{\ell_i}=1}}(1-\chi(\sigma_{\mathfrak{l}_i})\theta^{-1}(\ell_i)\ell_i^{-1}T_{\ell_i})\prod_{\substack{\mathfrak{l}_i|\ell_i\\ e_{\ell_i}=2}}(1-\chi(\sigma_{\mathfrak{l}_i})\theta^{-1}(\ell_i)\ell_i^{-1}T_{\ell_i}+\chi^2(\sigma_{\mathfrak{l}_i})\ell_i^{-1}).$$

Taking the limit over r, we thus obtain a $\mathbb{T}(\mathfrak{a})^{\circ}$-linear map

$$\mathbb{T}(\mathfrak{a})^{\circ}\otimes_{(\mathbb{T}_{N(\Sigma)}^{N^-})_{\mathfrak{n}}}(\mathfrak{J}_{N(\Sigma)^+})_{\mathfrak{n}} \to \mathbb{T}(\mathfrak{a})^{\circ}\otimes_{(\mathbb{T}_{N(\mathfrak{a})}^{N^-})_{\mathfrak{m}}}(\mathfrak{J}_{N(\mathfrak{a})^+})_{\mathfrak{m}} \tag{13}$$

having an effect on the corresponding measures as stated in Theorem 3.10. Hence to conclude the proof it remains to show that (13) is an isomorphism.

By Theorem 3.5, both the source and the target of this map are free of rank one over $\mathbb{T}(\mathfrak{a})^{\circ}$, and as in [Emerton et al. 2006, p. 559] (using [Hida 1988, Theorem 9.4]), one is reduced to showing the injectivity of the dual map modulo p:

$$\begin{aligned} H^0(\widetilde{X}_{N(\mathfrak{a})^+,1};\mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{m}] &\to (\mathbb{T}_{N(\mathfrak{a})}^{N^-}/\mathfrak{m})\otimes_{\mathbb{T}_{N(\Sigma)}^{N^-}/\mathfrak{n}}(H^0(\widetilde{X}_{N(\mathfrak{a})^+,1};\mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{m}']) \\ &\to (\mathbb{T}_{N(\mathfrak{a})}^{N^-}/\mathfrak{m})\otimes_{\mathbb{T}_{N(\Sigma)}^{N^-}/\mathfrak{n}}(H^0(\widetilde{X}_{N(\Sigma)^+,1};\mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{m}']) \\ &\to (\mathbb{T}_{N(\mathfrak{a})}^{N^-}/\mathfrak{m})\otimes_{\mathbb{T}_{N(\Sigma)}^{N^-}/\mathfrak{n}}(H^0(\widetilde{X}_{N(\Sigma)^+,1};\mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{n}]); \end{aligned} \tag{14}$$

or equivalently (by a version of [Emerton et al. 2006, Lemma 3.8.1]), to showing that the composite of the first two arrows in (14) is injective.

In turn, the latter injectivity follows from Lemma 3.11 below, where the notations are as follows:

- M^+ is any positive integer with $(M^+, pN^-) = 1$;
- $\ell \neq p$ is a prime;
- $n_\ell = 1$ or 2 according to whether or not ℓ divides M^+;
- $N^+ := \ell^{n_\ell} M^+$;

and

$$\epsilon_\ell^* : H^0(\widetilde{X}_{M^+,1}; \mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{m}] \to (\mathbb{T}^{N^-}_{M^+N^-}/\mathfrak{m}) \otimes_{\mathbb{T}'_{N^+N^-}/\mathfrak{m}'} (H^0(\widetilde{X}_{N^+,1}; \mathbb{F}_p)^{\mathrm{ord}}[\mathfrak{m}']) \quad (15)$$

is the map defined by

$$\epsilon_\ell^* := \begin{cases} B^*_{\ell,1} - B^*_{\ell,\ell}\ell^{-1}T_\ell & \text{if } n_\ell = 1; \\ B^*_{\ell^2,1} - B^*_{\ell^2,\ell}\ell^{-1}T_\ell + B^*_{\ell^2,\ell^2}\ell^{-1}\langle \ell \rangle_{N(\mathfrak{a})p} & \text{if } n_\ell = 2. \end{cases}$$

Lemma 3.11. *The map* (15) *is injective.*

Proof. As in the proof of the analogous result [Emerton et al. 2006, Lemma 3.8.2] in the modular curve case, it suffices to show the injectivity of the map

$$(H^0(\widetilde{X}_{M^+,1}; \mathbb{F})^{\mathrm{ord}}[\mathfrak{m}_{\mathbb{F}}])^{n_\ell+1} \xrightarrow{\beta_\ell} H^0(\widetilde{X}_{N^+,1}; \mathbb{F})^{\mathrm{ord}}[\mathfrak{m}'_{\mathbb{F}}]$$

defined by

$$\beta_\ell := \begin{cases} B^*_{\ell,1}\pi_1 + B^*_{\ell,\ell}\pi_2 & \text{if } n_\ell = 1; \\ B^*_{\ell^2,1}\pi_1 + B^*_{\ell^2,\ell}\pi_2 + B^*_{\ell^2,\ell^2}\pi_3 & \text{if } n_\ell = 2. \end{cases}$$

But in our quaternionic setting the proof of this injectivity follows from [Skinner and Wiles 1999, Lemma 3.26] for $n_\ell = 1$ and [*loc.cit.*, Lemma 3.28] for $n_\ell = 2$. □

Applying inductively Lemma 3.11 to the primes in Σ, the proof of Theorem 3.10 follows. □

3G. ***Analytic Iwasawa invariants.*** Upon the choice of an isomorphism

$$\mathbb{Z}_p[\![\Gamma]\!] \simeq \mathbb{Z}_p[\![T]\!]$$

we may regard the p-adic L-functions $L_\Sigma(\bar{\rho}, \mathfrak{a})$ and $L(\bar{\rho}, \mathfrak{a})$, as well as the Euler factor $E_\Sigma(\bar{\rho}, \mathfrak{a})$, as elements in $\mathbb{T}(\mathfrak{a})^\circ[\![T]\!]$. In this section we apply the main result of the preceding section to study the variation of the Iwasawa invariants attached to the anticyclotomic p-adic L-functions of p-ordinary modular forms.

For any power series $f(T) \in R[\![T]\!]$ with coefficients in a ring R, the *content* of $f(T)$ is defined to be the ideal $I(f(T)) \subseteq R$ generated by the coefficients of $f(T)$. If $\wp$ is a height one prime of $\mathbb{T}_\Sigma$ belonging to the branch $\mathbb{T}(\mathfrak{a})$ (in the sense that $\mathfrak{a}$ is the unique minimal prime of $\mathbb{T}_\Sigma$ contained in $\wp$), we denote by $L(\bar{\rho}, \mathfrak{a})(\wp)$ the element of $\mathcal{O}_\wp[\![\Gamma]\!]$ obtained from $L(\bar{\rho}, \mathfrak{a})$ via reduction modulo $\wp$. In particular, we note that $L(\bar{\rho}, \mathfrak{a})(\wp)$ has unit content if and only if its μ-invariant vanishes.

Theorem 3.12. *The following are equivalent*:

(1) $\mu(L(\bar{\rho},\mathfrak{a})(\wp))=0$ *for some newform* $f_\wp$ *in* $\mathcal{H}^-(\bar{\rho})$;

(2) $\mu(L(\bar{\rho},\mathfrak{a})(\wp))=0$ *for every newform* $f_\wp$ *in* $\mathcal{H}^-(\bar{\rho})$;

(3) $L(\bar{\rho},\mathfrak{a})$ *has unit content for some irreducible component* $\mathbb{T}(\mathfrak{a})$ *of* $\mathcal{H}^-(\bar{\rho})$;

(4) $L(\bar{\rho},\mathfrak{a})$ *has unit content for every irreducible component* $\mathbb{T}(\mathfrak{a})$ *of* $\mathcal{H}^-(\bar{\rho})$.

Proof. The argument in [Emerton et al. 2006, Theorem 3.7.5] applies verbatim, replacing the appeal to [*loc.cit.*, Corollary 3.6.3] by our Theorem 3.10 above. □

When any of the conditions in Theorem 3.12 hold, we shall write

$$\mu^{\mathrm{an}}(\bar{\rho})=0.$$

For a power series $f(T)$ with unit content and coefficients in a local ring R, the λ-invariant $\lambda(f(T))$ is defined to be the smallest degree in which $f(T)$ has a unit coefficient.

Theorem 3.13. *Assume that* $\mu^{\mathrm{an}}(\bar{\rho})=0$.

(1) *Let* $\mathbb{T}(\mathfrak{a})$ *be an irreducible component of* $\mathcal{H}^-(\bar{\rho})$. *As* $\wp$ *varies over the arithmetic primes of* $\mathbb{T}(\mathfrak{a})$, *the* λ*-invariant* $\lambda(L(\bar{\rho},\mathfrak{a})(\wp))$ *takes on a constant value, denoted* $\lambda^{\mathrm{an}}(\bar{\rho},\mathfrak{a})$.

(2) *For any two irreducible components* $\mathbb{T}(\mathfrak{a}_1)$, $\mathbb{T}(\mathfrak{a}_2)$ *of* $\mathcal{H}^-(\bar{\rho})$, *we have that*

$$\lambda^{\mathrm{an}}(\bar{\rho},\mathfrak{a}_1)-\lambda^{\mathrm{an}}(\bar{\rho},\mathfrak{a}_2)=\sum_{\ell\neq p}e_\ell(\mathfrak{a}_2)-e_\ell(\mathfrak{a}_1),$$

where $e_\ell(\mathfrak{a})=\lambda(E_\ell(\mathfrak{a}))$.

Proof. The first part follows immediately from the definitions. For the second part, the argument in [Emerton et al. 2006, Theorem 3.7.7] applies verbatim, replacing their appeal to [*loc.cit.*, Cor. 3.6.3] by our Theorem 3.10 above. □

By Theorem 3.12 and Theorem 3.13, the Iwasawa invariants of $L(\bar{\rho},\mathfrak{a})(\wp)$ are well behaved as $\wp$ varies. However, for the applications of these results to the Iwasawa main conjecture it is of course necessary to relate $L(\bar{\rho},\mathfrak{a})(\wp)$ to p-adic L-functions defined by the interpolation of special values of L-functions. This question was addressed in [Castella and Longo 2016], as we now recall.

Theorem 3.14. *If* $\wp$ *is the arithmetic prime of* $\mathbb{T}(\mathfrak{a})$ *corresponding to a* p*-ordinary* p*-stabilized newform* $f_\wp$ *of weight* $k\geqslant 2$ *and trivial nebentypus, then*

$$L(\bar{\rho},\mathfrak{a})(\wp)=L_p(f_\wp/K),$$

where $L_p(f_\wp/K)$ *is the* p*-adic* L*-function of Chida and Hsieh* [2016]. *In particular, if* $\chi:\Gamma\to\mathbb{C}_p^\times$ *is the* p*-adic avatar of an anticyclotomic Hecke character of* K

of infinity type $(m, -m)$ *with* $-k/2 < m < k/2$, *then* $L(\bar{\rho}, \mathfrak{a})(\wp)$ *interpolates the central critical values*

$$\frac{L(f_\wp/K, \chi, k/2)}{\Omega_{f_\wp, N^-}}$$

as χ *varies*, *where* $\Omega_{f_\wp, N^-}$ *is the complex Gross period* (6).

Proof. This is a reformulation of the main result of [Castella and Longo 2016]. (Note that the constant $\lambda_\wp \in F_\wp^\times$ in [Castella and Longo 2016, Theorem. 4.6] is not needed here, since the specialization map of [*loc.cit.*, §3.1] is being replaced by the map $(\mathfrak{J}_{N^+})_{\mathfrak{m}} \to (\mathfrak{J}_{N^+,r,k})_{\mathfrak{m}_{r,k}}$ induced by the isomorphism (8), which preserves integrality.) □

Corollary 3.15. *Let* $f_1, f_2 \in \mathcal{H}^-(\bar{\rho})$ *be newforms with trivial nebentypus lying in the branches* $\mathbb{T}(\mathfrak{a}_1), \mathbb{T}(\mathfrak{a}_2)$, *respectively. Then* $\mu^{\mathrm{an}}(\bar{\rho}) = 0$ *and*

$$\lambda(L_p(f_1/K)) - \lambda(L_p(f_2/K)) = \sum_{\ell \neq p} e_\ell(\mathfrak{a}_2) - e_\ell(\mathfrak{a}_1),$$

where $e_\ell(\mathfrak{a}_j) = \lambda(E_\ell(\mathfrak{a}_j))$.

Proof. By [Chida and Hsieh 2016, Theorem. 5.7] (extending Vatsal's result [2003] to higher weights), if $f \in \mathcal{H}^-(\bar{\rho})$ has weight $k \leqslant p+1$ and trivial nebentypus, then $\mu(L_p(f/K)) = 0$. By Theorems 3.12 and 3.14, this implies $\mu^{\mathrm{an}}(\bar{\rho}) = 0$. The result thus follows from Theorem 3.13, using Theorem 3.14 again to replace $\lambda^{\mathrm{an}}(\bar{\rho}, \mathfrak{a}_j)$ by $\lambda(L_p(f_j/K))$. □

4. Anticyclotomic Selmer groups

We keep the notation of the previous sections. In particular, $\bar{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F})$ is an odd irreducible Galois representation satisfying Assumption (SU) and isomorphic to $\bar{\rho}_{f_0}$ for some newform f_0 of weight 2, $\mathcal{H}^-(\bar{\rho})$ is the associated Hida family, and Σ is a finite set of primes split in the imaginary quadratic field K.

For each $f \in \mathcal{H}^-(\bar{\rho})$, let V_f denote the self-dual Tate twist of the p-adic Galois representation associated to f, fix an $\mathcal{O}$-stable lattice $T_f \subseteq V_f$, and set $A_f := V_f/T_f$. Since f is p-ordinary, there is a unique one-dimensional $G_{\mathbb{Q}_p}$-invariant subspace $F_p^+ V_f \subseteq V_f$ where the inertia group at p acts via $\varepsilon_{\mathrm{cyc}}^{k/2}\psi$, with ψ a finite order character. Let $F_p^+ A_f$ be the image of $F_p^+ V_f$ in A_f, and as recalled in the Introduction define the *minimal Selmer group* of f by

$$\mathrm{Sel}(K_\infty, f) := \ker\left\{ H^1(K_\infty, A_f) \to \prod_{w \nmid p} H^1(K_{\infty,w}, A_f) \times \prod_{w \mid p} H^1(K_{\infty,w}, F_p^- A_f) \right\},$$

where w runs over the places of K_∞ and we set $F_p^- A_f := A_f / F_p^+ A_f$.

It is well known that $\mathrm{Sel}(K_\infty, f)$ is cofinitely generated over Λ. When it is also Λ-cotorsion, we define the μ-invariant $\mu(\mathrm{Sel}(K_\infty, f))$ (resp. λ-invariant

$\lambda(\mathrm{Sel}(K_\infty, f))$) to the largest power of ϖ dividing (resp. the number of zeros of) the characteristic power series of the Pontryagin dual of $\mathrm{Sel}(K_\infty, f)$.

A distinguishing feature of the anticyclotomic setting (in comparison with cyclotomic Iwasawa theory) is the presence of primes which split infinitely in the corresponding $\mathbb{Z}_p$-extension. Indeed, being inert in K, all primes $\ell \mid N^-$ are infinitely split in K_∞/K. As a result, the above Selmer group differs in general from the *Greenberg Selmer group* of f, defined as

$$\mathfrak{Sel}(K_\infty, f) := \ker\left\{H^1(K_\infty, A_f) \to \prod_{w\nmid p} H^1(I_{\infty,w}, A_f) \times \prod_{w\mid p} H^1(K_{\infty,w}, F_p^- A_f)\right\},$$

where $I_{\infty,w} \subseteq G_{K_\infty}$ denotes the inertia group at w.

If S is a finite set of primes in K, we let $\mathrm{Sel}^S(K_\infty, f)$ and $\mathfrak{Sel}^S(K_\infty, f)$ be the "S-primitive" Selmer groups defined as above by omitting the local conditions at the primes in S (except those above p, when any such prime is in S). Moreover, if S consists of the primes dividing a rational integer M, we replace the superscript S by M in the above notation.

Immediately from the definitions, we see that there is as exact sequence

$$0 \to \mathrm{Sel}(K_\infty, f) \to \mathfrak{Sel}(K_\infty, f) \to \prod_{\ell \mid N^-} \mathcal{H}_\ell^{\mathrm{un}}, \tag{16}$$

where

$$\mathcal{H}_\ell^{\mathrm{un}} := \ker\left\{\prod_{w\mid\ell} H^1(K_{\infty,w}, A_f) \to \prod_{w\mid\ell} H^1(I_{\infty,w}, A_f)\right\}$$

is the set of unramified cocycles. In [Pollack and Weston 2011, §§3, 5], Pollack and Weston carried out a careful analysis of the difference between $\mathrm{Sel}(K_\infty, f)$ and $\mathfrak{Sel}(K_\infty, f)$. Even though [loc. cit.] is mostly concerned with cases in which f is of weight 2, many of their arguments apply more generally. In fact, the next result follows essentially from their work.

Theorem 4.1. *Assume that $\bar\rho$ satisfies Hypotheses (SU). Then the following are equivalent*:

(1) $\mathrm{Sel}(K_\infty, f_0)$ *is Λ-cotorsion with μ-invariant zero for some newform* $f_0 \in \mathcal{H}^-(\bar\rho)$;

(2) $\mathrm{Sel}(K_\infty, f)$ *is Λ-cotorsion with μ-invariant zero for all newforms* $f \in \mathcal{H}^-(\bar\rho)$;

(3) $\mathfrak{Sel}(K_\infty, f)$ *is Λ-cotorsion with μ-invariant zero for all newforms* $f \in \mathcal{H}^-(\bar\rho)$.

Moreover, *in that case* $\mathrm{Sel}(K_\infty, f) \simeq \mathfrak{Sel}(K_\infty, f)$.

Proof. Assume f_0 is a newform in $\mathcal{H}^-(\bar\rho)$ for which $\mathrm{Sel}(K_\infty, f_0)$ is Λ-cotorsion with μ-invariant zero, and set $N^+ := N(\Sigma)/N^-$. By [Pollack and Weston 2011, Proposition 5.1], we then have the exact sequences

$$0 \to \mathrm{Sel}(K_\infty, f_0) \to \mathrm{Sel}^{N^+}(K_\infty, f_0) \to \prod_{\ell \mid N^+} \mathcal{H}_\ell \to 0; \tag{17}$$

$$0 \to \mathfrak{Sel}(K_\infty, f_0) \to \mathfrak{Sel}^{N^+}(K_\infty, f_0) \to \prod_{\ell \mid N^+} \mathcal{H}_\ell \to 0, \tag{18}$$

where $\mathcal{H}_\ell$ is the product of $H^1(K_{\infty,w}, A_{f_0})$ over the places $w \mid \ell$ in K_∞. Since every prime $\ell \mid N^+$ splits in K (see Remark 3.9), the Λ-cotorsionness and the vanishing of the μ-invariant of $\mathcal{H}_\ell$ can be deduced from [Greenberg and Vatsal 2000, Proposition 2.4]. Since $\mathrm{Sel}(K_\infty, f_0)[\varpi]$ is finite by assumption, it thus follows from (17) that $\mathrm{Sel}^{N^+}(K_\infty, f_0)[\varpi]$ is finite. Combined with (16) and [Pollack and Weston 2011, Corollary 5.2], the same argument using (18) shows that then $\mathfrak{Sel}^{N^+}(K_\infty, f_0)[\varpi]$ is also finite.

On the other hand, following the arguments in the proof [Pollack and Weston 2011, Proposition 3.6] we see that for any $f \in \mathcal{H}(\bar{\rho})$ we have the isomorphisms

$$\mathrm{Sel}^{N^+}(K_\infty, \bar{\rho}) \simeq \mathrm{Sel}^{N^+}(K_\infty, f)[\varpi]; \quad \mathfrak{Sel}^{N^+}(K_\infty, \bar{\rho}) \simeq \mathfrak{Sel}^{N^+}(K_\infty, f)[\varpi].$$

As a result, the argument in the previous paragraph implies that, for any newform $f \in \mathcal{H}^-(\bar{\rho})$, both $\mathrm{Sel}^{N^+}(K_\infty, f)[\varpi]$ and $\mathfrak{Sel}^{N^+}(K_\infty, f)[\varpi]$ are finite, from where (using (17) and (18) with f in place of f_0) the Λ-cotorsionness and the vanishing of both the μ-invariant of $\mathrm{Sel}(K_\infty, f)$ and of $\mathfrak{Sel}(K_\infty, f)$ follows. In view of (16) and [Pollack and Weston 2011, Lemma 3.4], the result follows. □

Let w be a prime of K_∞ above $\ell \neq p$ and denote by $G_w \subseteq G_{K_\infty}$ its decomposition group. Let $\mathbb{T}(\mathfrak{a})$ be the irreducible component of $\mathbb{T}_\Sigma$ passing through f, and define

$$\delta_w(\mathfrak{a}) := \dim_{\mathbb{F}} A_f^{G_w}/\varpi.$$

(Note that this is well defined by [Emerton et al. 2006, Lemma 4.3.1].) Assume $\ell = \mathfrak{l}\bar{\mathfrak{l}}$ splits in K and put

$$\delta_\ell(\mathfrak{a}) := \sum_{w \mid \ell} \delta_w(\mathfrak{a}), \tag{19}$$

where the sum is over the (finitely many) primes w of K_∞ above ℓ.

In view of Theorem 4.1, we write $\mu^{\mathrm{alg}}(\bar{\rho}) = 0$ whenever any of the μ-invariants appearing in that result vanish. In that case, for any newform f in $\mathcal{H}^-(\bar{\rho})$ we may consider the λ-invariants $\lambda(\mathrm{Sel}(K_\infty, f)) = \lambda(\mathfrak{Sel}(K_\infty, f))$.

Theorem 4.2. *Let $\bar{\rho}$ and Σ be as above, and assume that $\mu^{\mathrm{alg}}(\bar{\rho}) = 0$. If f_1 and f_2 are any two newforms in $\mathcal{H}^-(\bar{\rho})$ lying in the branches $\mathbb{T}(\mathfrak{a}_1)$ and $\mathbb{T}(\mathfrak{a}_2)$, respectively, then*

$$\lambda(\mathrm{Sel}(K_\infty, f_1)) - \lambda(\mathrm{Sel}(K_\infty, f_2)) = \sum_{\ell \neq p} \delta_\ell(\mathfrak{a}_1) - \delta_\ell(\mathfrak{a}_2).$$

Proof. Since we have the divisibilities $N^-|N(\mathfrak{a}_i)|N(\Sigma)$ with the quotient $N(\Sigma)/N^-$ only divisible by primes that are split in K, the arguments of [Emerton et al. 2006, §4] apply verbatim (cf., [Pollack and Weston 2011, Theorem 7.1]). $\square$

5. Applications to the main conjecture

5A. ***Variation of anticyclotomic Iwasawa invariants.*** Recall the definition of the analytic invariant $e_\ell(\mathfrak{a}) = \lambda(E_\ell(\mathfrak{a}))$, where $E_\ell(\mathfrak{a})$ is the Euler factor from Section 3F, and of the algebraic invariant $\delta_\ell(\mathfrak{a})$ introduced in (19).

Lemma 5.1. *Let $\mathfrak{a}_1$, $\mathfrak{a}_2$ be minimal primes of $\mathbb{T}_\Sigma$. For any prime $\ell \neq p$ split in K,*

$$\delta_\ell(\mathfrak{a}_1) - \delta_\ell(\mathfrak{a}_2) = e_\ell(\mathfrak{a}_2) - e_\ell(\mathfrak{a}_1).$$

Proof. Let $\mathfrak{a}$ be a minimal prime of $\mathbb{T}_\Sigma$, let f be a newform in the branch $\mathbb{T}(\mathfrak{a})$, and let $\wp_f \subseteq \mathfrak{a}$ be the corresponding height one prime. Since $\ell = \mathfrak{l}\bar{\mathfrak{l}}$ splits in K, we have

$$\bigoplus_{w|\ell} H^1(K_{\infty,w}, A_f) = \Big(\bigoplus_{w|\mathfrak{l}} H^1(K_{\infty,w}, A_f)\Big) \oplus \Big(\bigoplus_{w|\bar{\mathfrak{l}}} H^1(K_{\infty,w}, A_f)\Big)$$

and [Greenberg and Vatsal 2000, Proposition 2.4] immediately implies that

$$Ch_\Lambda\Big(\bigoplus_{w|\ell} H^1(K_{\infty,w}, A_f)^\vee\Big) = E_\ell(f, \ell^{-1}\gamma_{\mathfrak{l}}) \cdot E_\ell(f, \ell^{-1}\gamma_{\bar{\mathfrak{l}}}),$$

where $E_\ell(f, \ell^{-1}\gamma_{\mathfrak{l}}) \cdot E_\ell(f, \ell^{-1}\gamma_{\bar{\mathfrak{l}}})$ is the specialization of $E_\ell(\mathfrak{a})$ at $\wp_f$. The result thus follows from [Emerton et al. 2006, Lemma 5.1.5]. $\square$

Theorem 5.2. *Suppose that $\bar{\rho}$ satisfies Assumption (SU). If for some newform $f_0 \in \mathcal{H}^-(\bar{\rho})$ we have the equalities*

$$\mu(\mathrm{Sel}(K_\infty, f_0)) = \mu(L_p(f_0/K)) = 0 \quad \textit{and} \quad \lambda(\mathrm{Sel}(K_\infty, f_0)) = \lambda(L_p(f_0/K)),$$

then the equalities

$$\mu(\mathrm{Sel}(K_\infty, f)) = \mu(L_p(f/K)) = 0 \quad \textit{and} \quad \lambda(\mathrm{Sel}(K_\infty, f)) = \lambda(L_p(f/K))$$

hold for all newforms $f \in \mathcal{H}^-(\bar{\rho})$.

Proof. Let f be any newform in $\mathcal{H}^-(\bar{\rho})$. Since the algebraic and analytic μ-invariants of f_0 both vanish, the vanishing of $\mu(\mathrm{Sel}(K_\infty, f))$ and $\mu(L_p(f/K))$ follows from Theorems 4.1 and 3.12, respectively. On the other hand, combining Theorems 3.13 and 4.2, and Lemma 5.1, we see that

$$\lambda(\mathrm{Sel}(K_\infty, f)) - \lambda(\mathrm{Sel}(K_\infty, f_0)) = \lambda(L_p(f/K)) - \lambda(L_p(f_0/K)),$$

and hence the equality $\lambda(\mathrm{Sel}(K_\infty, f_0)) = \lambda(L_p(f_0/K))$ implies the same equality for f. $\square$

5B. ***Applications to the main conjecture.*** As an immediate consequence of Weierstrass preparation theorem, Theorem 5.2 together with one the divisibilities predicted by the anticyclotomic main conjecture implies the full anticyclotomic main conjecture.

Theorem 5.3 (Skinner–Urban). *Let $f \in S_k(\Gamma_0(N))$ be a newform of weight $k \equiv 2 \pmod{p-1}$ and trivial nebentypus. Suppose that $\bar{\rho}_f$ satisfies Assumption (SU) and that p splits in K. Then*

$$(L_p(f/K)) \supseteq Ch_\Lambda(\mathrm{Sel}(K_\infty, f)^\vee).$$

Proof. This follows from specializing the divisibility in [Skinner and Urban 2014, Theorem 3.26] to the anticyclotomic line. Indeed, let $\boldsymbol{f} = \sum_{n \geqslant 1} \boldsymbol{a}_n(\boldsymbol{f}) q^n \in \mathbb{I}[\![q]\!]$ be the Λ-adic form with coefficients in $\mathbb{I} := \mathbb{T}(\mathfrak{a})^\circ$ associated with the branch of the Hida family containing f, let Σ be a finite set of primes as in Section 3E, let $\Sigma' \supseteq \Sigma$ be a finite set of primes of K containing Σ and all primes dividing $pN(\mathfrak{a})D_K$, and assume that Σ' contains at least one prime $\ell \neq p$ that splits in K. Under these assumptions, in [Skinner and Urban 2014, Theorem 3.26] it is shown that

$$(\mathfrak{L}_p^{\Sigma'}(\boldsymbol{f}/K)) \supseteq Ch_{\Lambda_{\boldsymbol{f}}(L_\infty)}(\mathfrak{Sel}^{\Sigma'}(L_\infty, A_{\boldsymbol{f}})^\vee), \tag{20}$$

where $L_\infty = K_\infty K_{\mathrm{cyc}}$ is the $\mathbb{Z}_p^2$-extension of K, $\Lambda_{\boldsymbol{f}}(L_\infty)$ is the three-variable Iwasawa algebra $\mathbb{I}[\![\mathrm{Gal}(L_\infty/K)]\!]$, and $\mathfrak{L}_p^{\Sigma'}(\boldsymbol{f}/K)$ and $\mathfrak{Sel}^{\Sigma'}(L_\infty, A_{\boldsymbol{f}})$ are the "Σ'-primitive" p-adic L-function and Selmer group defined in [Skinner and Urban 2014, §3.4.5] and [Skinner and Urban 2014, §§3.1.3, 3.1.10], respectively.

Recall the character $\Theta : G_{\mathbb{Q}} \to \mathbb{Z}_p[\![1+p\mathbb{Z}_p]\!]^\times$ from Section 2C, regarded as a character on $\mathrm{Gal}(L_\infty/K)$, and let

$$\mathrm{Tw}_{\Theta^{-1}} : \Lambda_{\boldsymbol{f}}(L_\infty) \to \Lambda_{\boldsymbol{f}}(L_\infty)$$

be the $\mathbb{I}$-linear isomorphism induced by $\mathrm{Tw}_{\Theta^{-1}}(g) = \Theta^{-1}(g)g$ for $g \in \mathrm{Gal}(L_\infty/K)$. Choose a topological generator $\gamma \in \mathrm{Gal}(K_{\mathrm{cyc}}/K)$, and expand

$$\mathrm{Tw}_{\Theta^{-1}}(\mathfrak{L}_p^{\Sigma'}(\boldsymbol{f}/K)) = \mathfrak{L}_{p,0}^{\Sigma'}(\boldsymbol{f}/K) + \mathfrak{L}_{p,1}^{\Sigma'}(\boldsymbol{f}/K)(\gamma - 1) + \cdots$$

with $\mathfrak{L}_{p,i}^{\Sigma'}(\boldsymbol{f}/K) \in \Lambda_{\boldsymbol{f}}(K_\infty) = \mathbb{I}[\![\Gamma]\!]$. In particular, note that $\mathfrak{L}_{p,0}^{\Sigma'}(\boldsymbol{f}/K)$ is the restriction of the twisted three-variable p-adic L-function $\mathrm{Tw}_{\Theta^{-1}}(\mathfrak{L}_p^{\Sigma'}(\boldsymbol{f}/K))$ to the "self-dual" plane.

Because of our assumptions on f, the Λ-adic form $\boldsymbol{f}$ has trivial tame character, and hence denoting by Frob_ℓ an arithmetic Frobenius at any prime $\ell \nmid N(\mathfrak{a})p$, the Galois representation

$$\rho(\mathfrak{a}) : G_{\mathbb{Q}} \to \mathrm{GL}(T_{\boldsymbol{f}}) \simeq \mathrm{GL}_2(\mathbb{T}(\mathfrak{a})^\circ)$$

considered in Section 1D (which is easily seen to agree with the twisted representation considered in [Skinner and Urban 2014, p. 37]) satisfies

$$\det(X-\mathrm{Frob}_\ell\,|T_{\boldsymbol{f}})=X^2-\boldsymbol{a}_\ell(\boldsymbol{f})X+\Theta^2(\ell)\ell.$$

The twist $T_{\boldsymbol{f}}^\dagger:=T_{\boldsymbol{f}}\otimes\Theta^{-1}$ is therefore self-dual. Thus combining [Rubin 2000, Lemma 6.1.2] with a straightforward variant of [Skinner and Urban 2014, Proposition 3.9] having $\mathrm{Gal}(K_\infty/K)$ in place of $\mathrm{Gal}(K_{\mathrm{cyc}}/K)$, we see that divisibility (20) implies that

$$(\mathfrak{L}_{p,0}^{\Sigma'}(\boldsymbol{f}/K))\supseteq Ch_{\Lambda_{\boldsymbol{f}}(K_\infty)}(\mathfrak{Sel}^{\Sigma'}(K_\infty,A_{\boldsymbol{f}}^\dagger)^\vee). \tag{21}$$

(Here, as above, $A_{\boldsymbol{f}}$ denotes the Pontryagin dual $T_{\boldsymbol{f}}\otimes_{\mathbb{I}}\mathrm{Hom}_{\mathrm{cts}}(\mathbb{I},\mathbb{Q}_p/\mathbb{Z}_p)$, and $A_{\boldsymbol{f}}^\dagger$ is the corresponding twist.) We next claim that, setting $\Sigma'':=\Sigma'\setminus\Sigma$, we have

$$(\mathfrak{L}_{p,0}^{\Sigma'}(\boldsymbol{f}/K))=\biggl(L_\Sigma(\bar\rho,\mathfrak{a})\cdot\prod_{v\in\Sigma'',v\nmid p}E_v(\mathfrak{a})\biggr), \tag{22}$$

where $L_\Sigma(\bar\rho,\mathfrak{a})$ is the two-variable p-adic L-function constructed in Section 3D, and if v lies over the rational prime ℓ, $E_v(\mathfrak{a})$ is the Euler factor given by

$$E_v(\mathfrak{a})=\det(\mathrm{Id}-\mathrm{Frob}_v\,X|(V_{\boldsymbol{f}}^\dagger)_{I_v})_{X=\ell^{-1}\,\mathrm{Frob}_v},$$

where $V_{\boldsymbol{f}}:=T_{\boldsymbol{f}}\otimes_{\mathbb{I}}\mathrm{Frac}(\mathbb{I})$, and Frob_v is an arithmetic Frobenius at v. (Note that for $\ell=\mathfrak{l}\bar{\mathfrak{l}}$ split in K, $E_{\mathfrak{l}}(\mathfrak{a})\cdot E_{\bar{\mathfrak{l}}}(\mathfrak{a})$ is simply the Euler factor (11).) Indeed, combined with Theorems 3.10 and 3.14, equality (22) specialized to any arithmetic prime $\wp\subseteq\mathbb{T}(\mathfrak{a})$ of weight 2 is shown in [Skinner and Urban 2014, (12.3)], from where the claim follows easily from the density of these primes. (See also [Pollack and Weston 2011, Theorem 6.8] for the comparison between the different periods involved in the two constructions, which differ by a p-adic unit under our assumptions.)

Finally, (21) and (22) combined with Theorem 3.10 and [Greenberg and Vatsal 2000, Propositions 2.3,8] imply that

$$(L(\bar\rho,\mathfrak{a}))\supseteq Ch_{\Lambda_{\boldsymbol{f}}(K_\infty)}(\mathfrak{Sel}(K_\infty,A_{\boldsymbol{f}}^\dagger)^\vee),$$

from where the result follows by specializing at $\wp_f$ using Theorem 3.14 and Theorem 4.1. □

In the opposite direction, we have the following result:

Theorem 5.4 (Bertolini–Darmon). *Let $f=\sum_{n=1}^\infty a_n(f)q^n$ be a p-ordinary newform of weight 2, level N, and trivial nebentypus. Suppose that $\bar\rho_f$ satisfies Assumption (CR) and that*

$$a_p(f)\not\equiv\pm1 \pmod p. \tag{PO}$$

Then

$$(L_p(f/K))\subseteq Ch_\Lambda(\mathrm{Sel}(K_\infty,f)^\vee).$$

Proof. This is the main result of [Bertolini and Darmon 2005], as extended by Pollack and Weston [Pollack and Weston 2011] to newforms of weight 2 not necessarily defined over $\mathbb{Q}$ and under the stated hypotheses (weaker that in [Bertolini and Darmon 2005]) on $\bar{\rho}_f$. See also [Kim et al. 2017] for a detailed discussion on the additional "nonanomalous" hypothesis (PO) on f. $\square$

Before we combine the previous two theorems with our main results in this paper, we note that condition (PO) in Theorem 5.4 can be phrased in terms of the Galois representation ρ_f associated to f. Indeed, let $f = \sum_{n=1}^{\infty} a_n(f)q^n$ be a p-ordinary newform as above, defined over a finite extension $F/\mathbb{Q}_p$ with ring of integers $\mathcal{O}$. Then

$$\rho_f|_{D_p} \simeq \begin{pmatrix} \varepsilon & * \\ 0 & \delta \end{pmatrix}$$

on a decomposition group $D_p \subseteq G_{\mathbb{Q}}$ at p, with $\delta : D_p \to \mathcal{O}^\times$ an unramified character sending Frob_p to the unit root α_p of $X^2 - a_p(f)X + p$. Since clearly $\alpha \equiv a_p(f) \pmod{p}$, we see that condition (PO) amounts to the requirement that

$$\delta(\mathrm{Frob}_p) \not\equiv \pm 1 \pmod{p}. \tag{PO}$$

Now we are finally in a position to prove our main application to the anticyclotomic Iwasawa main conjecture for p-ordinary newforms.

Corollary 5.5. *Suppose that $\bar{\rho}$ satisfies Assumptions* (SU) *and* (PO) *and that p splits in K, and let f be a newform in $\mathcal{H}^-(\bar{\rho})$ of weight $k \equiv 2 \pmod{p-1}$ and trivial nebentypus. Then the anticyclotomic Iwasawa main conjecture holds for f.*

Proof. After Theorems 5.2 and 5.3, to check the anticyclotomic main conjecture for *any* newform f as in the statement, it suffices to check the three equalities

$$\mu(\mathrm{Sel}(K_\infty, f_0)) = \mu(L_p(f_0/K)) = 0, \quad \lambda(\mathrm{Sel}(K_\infty, f_0)) = \lambda(L_p(f_0/K)) \tag{23}$$

hold for *some* $f_0 \in \mathcal{H}^-(\bar{\rho})$ of weight $k \equiv 2 \pmod{p-1}$ and trivial nebentypus.

Let $\mathbb{T}(\mathfrak{a})$ be the irreducible component of $\mathcal{H}^-(\bar{\rho})$ containing f, and let $f_0 \in S_2(\Gamma_0(Np))$ be the p-stabilized newform corresponding to an arithmetic prime $\wp \subseteq \mathbb{T}(\mathfrak{a})$ of weight 2 and trivial nebentypus. By Assumption (PO), the form f_0 is necessarily the p-stabilization of a p-ordinary newform $f_0^\sharp \in S_2(\Gamma_0(N))$ (see, e.g., [Howard 2007, Lemma 2.1.5]). From the combination of Theorems 5.3 and 5.4, the anticyclotomic Iwasawa main conjecture holds for $f_0^\sharp$, and since we clearly have

$$L_p(f_0/K) = L_p(f_0^\sharp/K) \quad \text{and} \quad \mathrm{Sel}(K_\infty, f_0) \simeq \mathrm{Sel}(K_\infty, f_0^\sharp)$$

(note that the latter isomorphism relies on the absolute irreducibility of $\bar{\rho}$), the anticyclotomic Iwasawa main conjecture holds for f_0 as well. In particular, equalities (23) hold for this f_0, and the result follows. $\square$

Acknowledgements

During the preparation of this paper, Castella was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 682152); Kim was partially supported by an AMS–Simons Travel Grant; Longo was partially supported by PRIN 2010-11 "Arithmetic Algebraic Geometry and Number Theory" and by PRAT 2013 "Arithmetic of Varieties over Number Fields".

References

[Bertolini and Darmon 1996] M. Bertolini and H. Darmon, "Heegner points on Mumford–Tate curves", *Invent. Math.* **126**:3 (1996), 413–456. MR Zbl

[Bertolini and Darmon 2005] M. Bertolini and H. Darmon, "Iwasawa's main conjecture for elliptic curves over anticyclotomic $\mathbb{Z}_p$-extensions", *Ann. of Math.* (2) **162**:1 (2005), 1–64. MR Zbl

[Castella and Longo 2016] F. Castella and M. Longo, "Big Heegner points and special values of L-series", *Ann. Math. Qué.* **40**:2 (2016), 303–324. MR Zbl

[Chida and Hsieh 2015] M. Chida and M.-L. Hsieh, "On the anticyclotomic Iwasawa main conjecture for modular forms", *Compos. Math.* **151**:5 (2015), 863–897. MR Zbl

[Chida and Hsieh 2016] M. Chida and M.-L. Hsieh, "Special values of anticyclotomic L-functions for modular forms", *Journal für die reine und angewandte Mathematik (Crelles Journal)* (online publication January 2016).

[Diamond and Taylor 1994] F. Diamond and R. Taylor, "Nonoptimal levels of mod l modular representations", *Invent. Math.* **115**:3 (1994), 435–462. MR Zbl

[Emerton et al. 2006] M. Emerton, R. Pollack, and T. Weston, "Variation of Iwasawa invariants in Hida families", *Invent. Math.* **163**:3 (2006), 523–580. MR Zbl

[Greenberg 1989] R. Greenberg, "Iwasawa theory for p-adic representations", pp. 97–137 in *Algebraic number theory*, edited by J. Coates et al., Adv. Stud. Pure Math. **17**, Academic Press, Boston, 1989. MR Zbl

[Greenberg and Vatsal 2000] R. Greenberg and V. Vatsal, "On the Iwasawa invariants of elliptic curves", *Invent. Math.* **142**:1 (2000), 17–63. MR Zbl

[Hida 1986] H. Hida, "Galois representations into $\mathrm{GL}_2(\mathbf{Z}_p[[X]])$ attached to ordinary cusp forms", *Invent. Math.* **85**:3 (1986), 545–613. MR Zbl

[Hida 1988] H. Hida, "On p-adic Hecke algebras for GL_2 over totally real fields", *Ann. of Math.* (2) **128**:2 (1988), 295–384. MR Zbl

[Howard 2007] B. Howard, "Variation of Heegner points in Hida families", *Invent. Math.* **167**:1 (2007), 91–128. MR Zbl

[Khare and Wintenberger 2009] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture. I", *Invent. Math.* **178**:3 (2009), 485–504. MR Zbl

[Kim 2017] C.-H. Kim, "Anticyclotomic Iwasawa invariants and congruences of modular forms", *Asian J. Math.* **21**:3 (2017), 499–530. MR

[Kim et al. 2017] C.-H. Kim, R. Pollack, and T. Weston, "On the freeness of anticyclotomic Selmer groups of modular forms", *Int. J. Number Theory* **13**:6 (2017), 1443–1455. MR Zbl

[Longo and Vigni 2011] M. Longo and S. Vigni, "Quaternion algebras, Heegner points and the arithmetic of Hida families", *Manuscripta Math.* **135**:3-4 (2011), 273–328. MR Zbl

[Pollack and Weston 2011] R. Pollack and T. Weston, "On anticyclotomic μ-invariants of modular forms", *Compos. Math.* **147**:5 (2011), 1353–1381. MR Zbl

[Rubin 2000] K. Rubin, *Euler systems*, Annals of Mathematics Studies **147**, Princeton University Press, Princeton, NJ, 2000. Hermann Weyl Lectures. The Institute for Advanced Study. MR Zbl

[Skinner and Urban 2014] C. Skinner and E. Urban, "The Iwasawa main conjectures for GL_2", *Invent. Math.* **195**:1 (2014), 1–277. MR Zbl

[Skinner and Wiles 1999] C. M. Skinner and A. J. Wiles, "Residually reducible representations and modular forms", *Inst. Hautes Études Sci. Publ. Math.* 89 (1999), 5–126. MR Zbl

[Vatsal 2003] V. Vatsal, "Special values of anticyclotomic L-functions", *Duke Math. J.* **116**:2 (2003), 219–261. MR Zbl

Communicated by John Henry Coates
Received 2017-01-09 Revised 2017-09-05 Accepted 2017-10-23

fcabello@math.princeton.edu *Department of Mathematics, Princeton University, Princeton, NJ, United States*

chanho.math@gmail.com *School of Mathematics, Korea Institute for Advanced Study (KIAS), Seoul, South Korea*

mlongo@math.unipd.it *Dipartimento di Matematica, Università di Padova, Padova, Italy*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2369

Effective nonvanishing for Fano weighted complete intersections

Marco Pizzato, Taro Sano and Luca Tasin

We show that the Ambro–Kawamata nonvanishing conjecture holds true for a quasismooth WCI X which is Fano or Calabi–Yau, i.e., we prove that, if H is an ample Cartier divisor on X, then $|H|$ is not empty. If X is smooth, we further show that the general element of $|H|$ is smooth. We then verify the Ambro–Kawamata conjecture for any quasismooth weighted hypersurface. We also verify Fujita's freeness conjecture for a Gorenstein quasismooth weighted hypersurface.

For the proofs, we introduce the arithmetic notion of regular pairs and highlight some interesting connections with the Frobenius coin problem.

1. Introduction

Complete intersections in weighted projective spaces (WCIs for short) form a natural class of varieties which are particularly interesting from the point of view of higher dimensional algebraic geometry. We refer to [Dolgachev 1982], [Mori 1975] and [Dimca 1986] for a general treatment of these varieties.

Reid [1980; 1987] and Iano-Fletcher [2000] systematically investigated notable examples of WCIs and started their classification. Several results have since been obtained concerning boundedness and classification; see for example [Johnson and Kollár 2001; Chen et al. 2011; Ballico et al. 2013; Chen 2015; Przyjalkowski and Shramov 2016].

The main motivation of this paper is to study the following conjecture in the realm of WCIs, in particular for what concerns the case of Fano and Calabi–Yau varieties.

Conjecture 1.1 (Ambro–Kawamata). *Let (X, Δ) be a klt pair and H be an ample Cartier divisor on X such that $H - K_X - \Delta$ is ample. Then $|H| \neq \varnothing$.*

For an introduction to this conjecture, see [Ambro 1999] and in particular [Kawamata 2000] where the 2-dimensional case is proven. In the smooth setting, Ionescu [Lanteri et al. 1993, p. 321] and Beltrametti and Sommese [1995] proposed related conjectures.

MSC2010: primary 14M10; secondary 11D04, 14J45.
Keywords: weighted complete intersections, nonvanishing, Ambro–Kawamata conjecture.

The Ambro–Kawamata conjecture is known to be true in full generality only in dimensions 1 and 2. Several cases have been studied, especially in dimension 3; see for instance [Xie 2009; Broustet and Höring 2010; Höring 2012; Cao and Jiang 2016].

A *fundamental divisor* on a Fano variety X is an ample Cartier divisor H which is primitive and proportional to $-K_X$. In the classification of Fano varieties, it is important to investigate the properties of the general member of the linear system given by H; see for instance [Ambro 1999]. The second purpose of this note is to study this problem in the case of Fano and Calabi–Yau smooth WCIs.

The main result of this paper is the following.

Theorem 1.2. *Let $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a well-formed quasismooth weighted complete intersection which is not a linear cone and H be an ample Cartier divisor on X. Assume that X is Fano or Calabi–Yau. Then $|H| \neq \varnothing$.*

Moreover, if X is smooth, then the number of $a_i = 1$ is at least c and the general element of $|\mathcal{O}_X(1)|$ is smooth.

For a smooth Fano WCI, it was already proved in [Przyjalkowski and Shramov 2016, Lemma 3.3] that at least two weights are 1, which implies the nonvanishing for a smooth Fano WCI. In addition, it is easy to prove Conjecture 1.1 for any smooth WCI of codimension 1 and 2, see Remark 4.9.

It is particularly interesting that, in the smooth case, we can prove the smoothness of the general member of the fundamental linear system (Corollary 5.3(ii)).

Theorem 1.2 is a direct consequence of Corollaries 5.3 and 5.13. In particular, in Corollary 5.3, we show that if $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ is a smooth well-formed Fano WCI which is not a linear cone, then the number of i for which $a_i = 1$ is at least $c + 1$. By using this, we can then show that the general element of $|\mathcal{O}_X(1)|$ is quasismooth (from which smoothness follows easily). One can not expect a similar statement for a general member of the fundamental linear system of a singular quasismooth WCI, as Example 5.8 shows. We also give a description of the base locus of $|\mathcal{O}_X(1)|$ in Remark 5.5 and an example whose base locus $\operatorname{Bs}|\mathcal{O}_X(1)|$ is singular and not quasismooth in Example 5.6.

In [Przyjalkowski and Shramov 2017, Corollary 4.2], the authors show that for a smooth well-formed Fano WCI X the number of a_i equal to 1 is at least $I(X) = \sum a_i - \sum d_j$ when $c \leq 2$ and write that they expect this to hold for any codimension. As a consequence of Proposition 5.2, we can confirm this expectation; see Corollary 5.11.

In the case of hypersurfaces, we can prove the following stronger result, which is the combination of Propositions 6.2 and 6.3:

Theorem 1.3. *Let $X = X_d \subset \mathbb{P} = \mathbb{P}(a_0, \dots, a_n)$ be a well-formed quasismooth hypersurface of degree d which is not a linear cone.*

(1) *If H is an ample Cartier divisor on X such that $H - K_X$ is ample, then $|H|$ is not empty.*

(2) *If X is Gorenstein and H is an ample Cartier divisor, then $K_X + mH$ is globally generated for any $m \geq n$.*

The second part of the statement is known as Fujita's freeness conjecture and it has been proven in the smooth setting up to dimension 5; see [Reider 1988; Ein and Lazarsfeld 1993; Kawamata 1997; Ye and Zhu 2015].

The methods. The above theorems are obtained by studying the arithmetic properties of quasismooth WCIs. More precisely, in Section 3, we prove a criterion (Proposition 3.1) for a WCI to be quasismooth, which generalizes Iano-Fletcher's criterion in codimension 1 and 2 (see [Iano-Fletcher 2000, Section 8]). We then exploit some arithmetic consequences of quasismoothness. In particular, Proposition 3.6 motivates the introduction of an *h-regular pair* (see Definition 4.1) which turns out to be a key tool in our treatment.

Given a positive integer h, a pair $(d;a) = (d_1,\dots,d_c;a_0,\dots,a_n) \in \mathbb{N}^c \times \mathbb{N}^{n+1}$ is said to be h-regular if for any $I = \{i_1,\dots,i_k\} \subset \{0,\dots,n\}$ such that $a_I := \gcd(a_{i_1},\dots,a_{i_k}) > 1$, either $a_I \mid h$ or there are distinct integers $p_1,\dots,p_k$ such that

$$a_I \mid d_{p_1}, \dots, d_{p_k}.$$

Set $\delta(d;a) := \sum_{j=1}^{c} d_j - \sum_{i=0}^{n} a_i$. By Proposition 3.6, any quasismooth (wellformed) WCI $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ gives rise to an h-regular pair $(d;a) = (d_1, \dots, d_c; a_0, \dots, a_n)$, where h is the smallest positive integer for which $\mathcal{O}_X(h)$ is Cartier. Remembering that $K_X = \mathcal{O}_X(\delta)$, the nonvanishing for a Fano or Calabi–Yau WCI follows from Proposition 5.12, which says that, if $(d;a)$ is h-regular such that $a_i \neq d_j$ and $a_i \nmid h$ for any i, j, then $\delta(d;a) > 0$. A more accurate statement (Corollary 5.3) is needed to prove that, if X is smooth, then the general element of $|\mathcal{O}_X(1)|$ is also smooth.

We now spend some words for the case $h = 1$. In this case, the pair $(d;a)$ is simply called *regular*. A smooth WCI X gives rise to a regular pair $(d;a)$. The nonvanishing is then equivalent to prove that

$$\delta(d;a) \geq G(a_0, \dots, a_n),$$

where $G(a_0, \dots, a_n)$ is the Frobenius number of $a_0, \dots, a_n$, i.e., the greatest integer which is not a nonnegative integral combination of $a_0, \dots, a_n$. In Conjecture 4.8, we speculate that $\delta(d;a) \geq G(a_0, \dots, a_n)$ for a regular pair $(d;a)$, under some natural assumptions. This would imply the Ambro–Kawamata conjecture for any smooth WCI.

We believe that this conjecture is interesting also from the arithmetic point of view, since it would give new bounds for the Frobenius number (see page 2382 for details).

2. Preliminaries and notation

In this section, we recall some basic facts about weighted complete intersections and fix our notation. See [Dolgachev 1982] or [Iano-Fletcher 2000] for further details.

Let $\mathbb{N}$ (resp. $\mathbb{N}_+$) be the set of nonnegative (resp. positive) integers. Let $a_0,\dots,a_n \in \mathbb{N}_+$. We define $\mathbb{P} := \mathbb{P}(a_0,\dots,a_n)$ to be the weighted projective space with weights $a_0,\dots,a_n$, i.e., $\mathbb{P} = \operatorname{Proj}\mathbb{C}[x_0,\dots,x_n]$, where x_i has weight a_i. We denote

$$\mathbb{P}(\underbrace{b_1,\dots,b_1}_{k_1},\dots,\underbrace{b_l,\dots,b_l}_{k_l})$$

by $\mathbb{P}(b_1^{(k_1)},\dots,b_l^{(k_l)})$ for short.

Note that if we start with $x_0,\dots,x_n$ to be affine coordinates on $\mathbb{A}^{n+1}$ and $\mathbb{C}^*$ acting on $\mathbb{A}^{n+1}$ via

$$\lambda\cdot(x_0,\dots,x_n) = (\lambda^{a_0}x_0,\dots,\lambda^{a_n}x_n)$$

for any $\lambda \in \mathbb{C}^*$, then $\mathbb{P}$ is just the quotient $(\mathbb{A}^{n+1}\setminus\{0\})/\mathbb{C}^*$.

We always assume that $\mathbb{P}$ is *well-formed*, i.e., the greatest common divisor of any n weights is 1. For any $I = \{i_1,\dots,i_k\} \subset \{0,\dots,n\}$, the stratum Π_I is defined as

$$\Pi_I := \{x_i = 0 : i \notin I\}.$$

The singular locus of $\mathbb{P}$ is the union of all strata Π_I for which $a_I := \gcd(a_i)_{i\in I} > 1$. Any point of the interior Π_I^0 of a stratum Π_I is locally isomorphic to a quotient singularity of type

$$\frac{1}{a_I}(a_0,\dots,\hat{a}_{i_1},\dots,\hat{a}_{i_k},\dots,a_n)\times\mathbb{C}^{k-1}.$$

Here, for $r \in \mathbb{N}_+$ and $a_1,\dots,a_n \in \mathbb{N}$ such that $\gcd(r,a_1,\dots,a_n) = 1$, a *quotient singularity* of type $1/r(a_1,\dots,a_n)$ means a quotient $\mathbb{C}^n/\mathbb{Z}_r$ by the action of a cyclic group $\mathbb{Z}_r$ of order r as $g\cdot z_i = \zeta_r^{a_i}z_i$ for $i = 1,\dots,n$, where $g \in \mathbb{Z}_r$ is a generator and ζ_r is an r-th primitive root of unity. We also denote by $\mathbb{C}^n/\mathbb{Z}_r(a_1,\dots,a_r)$ this quotient affine variety. Let $\pi : \mathbb{C}^n \to U := \mathbb{C}^n/\mathbb{Z}_r(a_1,\dots,a_r)$ be the quotient morphism. We have an eigendecomposition

$$\pi_*\mathcal{O}_{\mathbb{C}^n} = \bigoplus_{i=0}^{r-1}\mathcal{F}_i,$$

where $\mathcal{F}_i := \{f \in \mathcal{O}_{\mathbb{C}^n} \mid g \cdot f = \zeta_r^i f\}$ is the $\mathcal{O}_U$-submodule of $\pi_*\mathcal{O}_{\mathbb{C}^n}$ consisting of $\mathbb{Z}_r$-eigenfunctions of eigenvalue ζ_r^i. Note that $\mathcal{F}_i \simeq \mathcal{O}_U(D_f)$, where a divisor $D_f := (f=0)/\mathbb{Z}_r$ on U is defined by a function $f \in \mathcal{F}_i$.

Proposition 2.1. *The divisor class group of* $U := \mathbb{C}^n/\mathbb{Z}_r(a_1, \dots, a_r)$ *is*

$$\mathrm{Cl}\, U \simeq \mathbb{Z}_r \cdot \mathcal{F}_1. \tag{1}$$

Proof. We have an inclusion $\iota : \mathbb{Z}_r \cdot \mathcal{F}_1 \hookrightarrow \mathrm{Cl}\, U$. It is enough to show that this is surjective. Let $D \subset U$ be a prime divisor. Then $\pi^{-1}(D)$ is a divisor on $\mathbb{C}^n$ defined by some $\mathbb{Z}_r$-eigenfunction $f_D \in \mathcal{O}_{\mathbb{C}^n}$. Let $i(D) \in \mathbb{Z}_r$ be an element such that $g \cdot f_D = \zeta_r^{i(D)} f$. Then we can check that $\mathcal{O}_U(D) \simeq \mathcal{F}_{i(D)}$. Since $g^i \cdot \mathcal{F}_1 \simeq \mathcal{F}_i$ for $i \in \mathbb{Z}_r$, we see that ι is surjective.

We can also check the isomorphism by toric computation. Since $\mathbb{C}^n/\mathbb{Z}_r(a_1,\dots,a_n)$ is a toric variety, we can compute its class group by using the information of the cone and lattice. (cf., [Fulton 1993, p. 63, Proposition]) More precisely, it is the quotient $\mathbb{Z}^n/M$, where $M := \{(m_1,\dots,m_n) \in \mathbb{Z}^n \mid \sum_{i=1}^n m_i a_i \equiv 0 \mod r\}$. □

Definition 2.2. Let X be a (closed) subvariety of codimension c in $\mathbb{P}$. Then X is *well-formed* if

$$\mathrm{codim}_X(X \cap \mathrm{Sing}(\mathbb{P})) \geq 2.$$

Let $\pi : \mathbb{A}^{n+1} \setminus \{0\} \to \mathbb{P}$ be the natural projection. Then X is *quasismooth* if $\pi^{-1}(X)$ is smooth.

The variety X is said to be a *weighted complete intersection* (WCI for short) of multidegree $(d_1, \dots, d_c)$ if its weighted homogenous ideal in $\mathbb{C}[x_0, \dots, x_n]$ is generated by a regular sequence of homogenous polynomials $\{f_j\}$ such that $\deg f_j = d_j$ for $j = 1, \dots, c$. We denote by $X_{d_1,\dots,d_c}$ a general element of the family of WCIs of multidegree $(d_1, \dots, d_c)$.

Finally, $X_{d_1,\dots,d_c} \subset \mathbb{P}$ is said to be a *linear cone* if $d_j = a_i$ for some i and j.

Note that by [Dimca 1986, Proposition 8], if X is a well-formed quasismooth WCI, then

$$\mathrm{Sing}(X) = X \cap \mathrm{Sing}(\mathbb{P}).$$

Proposition 2.3. *If X is a quasismooth WCI of dimension* ≥ 3, *then its divisor class group is a free $\mathbb{Z}$-module generated by $\mathcal{O}_X(1)$, where $\mathcal{O}_X(1) := \mathcal{O}_{\mathbb{P}}(1)_{|X}$.* (*We freely mix the divisorial and the sheaf notation.*)

Proof. The proof is the same as [Corti et al. 2000, Lemma 3.5]. This follows from the parafactoriality of an l.c.i. local ring [Call and Lyubeznik 1994]. □

If $X \subset \mathbb{P}$ is a well-formed quasismooth WCI, then

$$\omega_X = K_X = \mathcal{O}_X\left(\sum_{j=1}^{c} d_j - \sum_{i=0}^{n} a_i\right),$$

see [Dolgachev 1982, Theorem 3.3.4]. We usually write $\delta := \sum_{j=1}^{c} d_j - \sum_{i=0}^{n} a_i$.

The following result shows that the dimension of the linear system $|\mathcal{O}_X(n)|$ can be computed by the weights of the coordinates.

Lemma 2.4 [Iano-Fletcher 2000, Lemma 7.1]. *Let $X \subset \mathbb{P}(a_0, \dots, a_n)$ be a well-formed quasismooth WCI. Let $A := k[x_0, \dots, x_n]/(f_1, \dots, f_c)$ be the homogeneous coordinate ring of X and A_k be the k-th graded part for $k \in \mathbb{Z}$. Then*

$$H^0(X, \mathcal{O}_X(k)) \simeq A_k.$$

Proof. See also [Dolgachev 1982, 3.4.3]. This follows since the homogeneous coordinate ring A is Cohen–Macaulay and $H^1_{\mathfrak{m}}(A) = 0$, where $\mathfrak{m} := (x_0, \dots, x_n)$ is the maximal ideal. □

3. Properties of quasismooth WCIs

In the following proposition, we give a necessary and sufficient condition for quasismoothness of a WCI.

Proposition 3.1. *Let $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n) =: \mathbb{P}$ be a quasismooth WCI which is not a linear cone. Let $x_0, \dots, x_n$ be the coordinates of $\mathbb{P}(a_0, \dots, a_n)$. Fix $I := \{i_1, \dots, i_k\} \subset \{0, \dots, n\}$ and let $\rho_I := \min\{c, k\}$. For $m = (m_1, \dots, m_k)$, let $x_I^m := \prod_{j=1}^{k} x_{i_j}^{m_j}$. For a finite set A, let $|A|$ be the number of its elements. Then one of the following holds.*

(Q1) *There exist distinct integers $p_1, \dots, p_{\rho_I} \in \{1, \dots, c\}$ and $M_1, \dots, M_{\rho_I} \in \mathbb{N}^k$ such that the monomial $x_I^{M_j}$ has the degree d_{p_j} for $j = 1, \dots, \rho_I$.*

(Q2) *There exist a permutation $p_1, \dots, p_c$ of $\{1, \dots, c\}$, an integer $l < \rho_I$, and integers $e_{\mu,j} \in \{0, \dots, n\} \setminus I$ for $\mu = 1, \dots, k-l$ and $j = l+1, \dots, c$ such that there are monomials $x_I^{M_j}$ of degree d_{p_j} for $j = 1, \dots, l$ and distinct $k-l$ monomials $\{x_{e_{\mu,j}} x_I^{M_{\mu,j}} : \mu = 1, \dots, k-l\}$ of degree d_{p_j} for each $j = l+1, \dots, c$ which satisfy the following: for any subset $J \subset \{l+1, \dots, c\}$, we have $|\{e_{\mu,j} : j \in J, \mu = 1, \dots, k-l\}| \geq k-l+|J|-1$.*

Conversely, if we have (Q1) *or* (Q2) *for all I, then a general WCI $X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ is quasismooth.*

Remark 3.2. This generalizes [Iano-Fletcher 2000, Theorem 8.7] in codimension 2 case. A weaker necessary condition for the quasismoothness is written in [Chen 2015, Proposition 2.3]. Although we shall not use the new part of Proposition 3.1 in the main part of this paper, we believe it is an interesting result on its own.

Proof. The framework of the proof is similar to that of [Iano-Fletcher 2000, Theorem 8.7].

Let $F_j := |\mathcal{O}_{\mathbb{P}}(d_j)|$ be the linear system of weighted homogeneous polynomials of degree d_j. For $j = 1, \dots, c$, let f_j be a general homogeneous polynomial of degree d_j such that $X = (f_1 = \cdots = f_c = 0) \subset \mathbb{P}(a_0, \dots, a_n)$. Let $C_X^* \subset \mathbb{A}^{n+1} \setminus \{0\}$ be the cone over X defined by the polynomials $f_1, \dots, f_c$ with the following diagram

$$\begin{array}{ccc} C_X^* & \hookrightarrow & \mathbb{A}^{n+1} \setminus \{0\} \\ \downarrow & & \downarrow \\ X & \hookrightarrow & \mathbb{P} \end{array}$$

Without loss of generality, we may assume $I = \{0, \dots, k-1\}$ in the statement. Let $\Pi := (x_k = \cdots = x_n = 0) \subset \mathbb{A}^{n+1}$ be the stratum corresponding to I and $\Pi^0 \subset \Pi$ be the open toric stratum. By expanding f_λ for $\lambda = 1, \dots, c$ in terms of $x_k, \dots, x_n$, we can write

$$f_\lambda = h_\lambda(x_0, \dots, x_{k-1}) + \sum_{i=k}^{n} x_i g_\lambda^i(x_0, \dots, x_{k-1}) + R_\lambda(x_0, \dots, x_n),$$

where $h_\lambda, g_\lambda^i \in \mathbb{C}[x_0, \dots, x_{k-1}]$ and $R_\lambda \in \mathbb{C}[x_0, \dots, x_n]$ satisfies $\deg_{x_k,\dots,x_n} R_\lambda \geq 2$.

Note that X is quasismooth if and only if C_X^* is smooth along all the coordinate strata. We shall show that C_X^* is smooth along Π^0 when either (Q1) or (Q2) holds for I. Let $\rho := \rho_I$ for short.

Suppose that (Q1) holds. Then $h_{p_1}, \dots, h_{p_\rho}$ are nonzero on Π^0. If some of h_{p_j} involves only one monomial, then we have $\Pi^0 \cap C_X^* = \varnothing$. So we may assume that each of $h_{p_1}, \dots, h_{p_\rho}$ involves at least 2 monomials. Thus we see that the linear systems $F_{d_{p_1}}, \dots, F_{d_{p_\rho}}$ do not have base locus on Π^0. By Bertini's theorem, we see that $(f_{p_1} = \cdots = f_{p_\rho} = 0) \subset \mathbb{A}^{n+1}$ is smooth along Π^0 when $k \geq c$. When $k < c$, we have $(f_{p_1} = \cdots = f_{p_\rho} = 0) \cap \Pi^0 = \varnothing$. Therefore C_X^* is nonsingular along Π^0.

Next suppose that (Q2) holds. By permutation, we may assume that $p_i = i$. Then $h_1, \dots, h_l$ are nonzero on Π^0. Hence the base locus of F_{d_λ} is disjoint with Π^0 for $\lambda = 1, \dots, l$. By Bertini's theorem, we see that $(f_1 = \cdots = f_l = 0)$ is nonsingular along Π^0. We may assume that the Jacobian of $(f_1 = \cdots = f_c = 0) \subset \mathbb{A}^{n+1}$ at $P \in \Pi^0$ is of the form

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_0} \cdots \frac{\partial f_1}{\partial x_{k-1}} & \\ \vdots \qquad \vdots & * \\ \frac{\partial f_l}{\partial x_0} \cdots \frac{\partial f_l}{\partial x_{k-1}} & \\ & g_{l+1}^k \cdots g_{l+1}^n \\ \mathbf{0} & \vdots \qquad \vdots \\ & g_c^k \cdots g_c^n \end{pmatrix} (P),$$

since we have $h_\lambda = 0$ for $\lambda = l+1, \dots, c$. Note that the block matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_0} \cdots \frac{\partial f_1}{\partial x_{k-1}} \\ \vdots \qquad \vdots \\ \frac{\partial f_l}{\partial x_0} \cdots \frac{\partial f_l}{\partial x_{k-1}} \end{pmatrix} (P)$$

has maximal rank l at $P \in \Pi^0$ since $(f_1 = \cdots = f_l = 0)$ is nonsingular along Π^0. Hence it is enough to show that the matrix

$$M_P := \begin{pmatrix} g^k_{l+1} \cdots g^n_{l+1} \\ \vdots \qquad \vdots \\ g^k_c \ \cdots \ g^n_c \end{pmatrix} (P) \tag{2}$$

has maximal rank $c - l$.

Note that there are at least $k - l$ elements of $K_\lambda := \{i \in \{k, \ldots, n\} : g^i_\lambda \neq 0\}$ for each $\lambda = l + 1, \ldots, c$. By $|K_\lambda| \geq k - l$, we see that each row vector of M_P is nonzero for $P \in \Pi^0$. Indeed, for each $\lambda' = l + 1, \ldots, c$, the intersection

$$\bigcap_{\lambda=1}^{l} (h_\lambda = 0) \cap \bigcap_{i=k}^{n} (g^i_{\lambda'} = 0) \cap \Pi^0$$

is contained in at least $k = l + (k - l)$ free linear systems on k-dimensional Π^0, and it is empty. Thus we may assume that $g^k_{l+1}(P) \neq 0$. We shall make elementary matrix operations on M_P to calculate the rank of M_P.

For $\lambda = l + 2, \ldots, c$, let

$$Z_\lambda(P) := \{Q \in (f_1 = \cdots = f_l = 0) \cap \Pi^0 : \\ g^k_{l+1}(P) g^i_\lambda(Q) - g^k_\lambda(P) g^i_{l+1}(Q) = 0 \, (i = k + 1, \ldots, n)\}.$$

Note that the first row M^1_P and the $(\lambda - l)$-th row $M^{\lambda-l}_P$ of M_P are linearly dependent if and only if $P \in Z_\lambda(P)$. By condition (Q2) for J with $|J| = 2$, there are at least $k - l$ nonzero elements of $G_\lambda(P) := \{g^k_{l+1}(P) g^i_\lambda - g^k_\lambda(P) g^i_{l+1} : i = k + 1, \ldots, n\}$ and they define $k - l$ free linear systems on Π^0. Hence we obtain $Z_\lambda(P) = \varnothing$ and the two rows M^1_P and $M^{\lambda-l}_P$ are linearly independent. Thus, by elementary operations on M_P, we obtain a matrix of the following form;

$$\begin{pmatrix} g^k_{l+1} & \cdots & \cdots & g^n_{l+1} \\ 0 & h^{k+1}_{l+2} & \cdots & h^n_{l+2} \\ \vdots & \vdots & & \vdots \\ 0 & h^{k+1}_c & \cdots & h^n_c \end{pmatrix} (P).$$

By column exchange operations, we may assume that $h^{k+1}_{l+2}(P) \neq 0$ and repeat the process to

$$M'_P := \begin{pmatrix} h^{k+1}_{l+2} & \cdots & h^n_{l+2} \\ \vdots & & \vdots \\ h^{k+1}_c & \cdots & h^n_c \end{pmatrix} (P).$$

Let $G'_\lambda(P) := \{h^{k+1}_{l+2}(P) h^i_\lambda - h^{k+1}_\lambda(P) h^i_{l+2} : i = k+2, \ldots, n\}$. By condition (Q2) for J with $|J| = 3$, there are at least $k - l$ nonzero elements of $G'_\lambda(P)$ and they define free linear systems on Π^0. By this, we again see that the first row and another row of M'_P are linearly independent.

After repeating these elementary operations, we obtain a matrix of the form

$$\begin{pmatrix} \alpha_{l+1} & & * & * \\ & \ddots & & \\ 0 & & \alpha_c & \end{pmatrix}$$

for some $\alpha_{l+1}, \ldots, \alpha_c \in \mathbb{C} \setminus \{0\}$ and see that the rank of M_P is $c-l$. Thus C_X^* is nonsingular at $P \in \Pi^0$.

Suppose that conditions (Q1) and (Q2) do not hold for some I. We shall show that X is not quasismooth. We may again assume that $I = \{0, \ldots, k-1\}$ and $\Pi = (x_k = \cdots = x_n = 0)$. Moreover, since (Q1) and (Q2) do not hold, we may assume that, for some $l < \rho_I$, we have $\Pi \not\subset (f_\lambda = 0)$ for $\lambda = 1, \ldots, l$ and $\Pi \subset (f_\lambda = 0)$ for $\lambda = l+1, \ldots, c$. Then the singular locus of C_X^* on Π^0 can be described as

$$Z := \{P \in (f_1 = \cdots = f_l = 0) \cap \Pi^0 : \operatorname{rk} M_P < c - l\},$$

where M_P is the matrix defined in (2). By the hypothesis, we may also assume that there exists $J \subset \{l+1, \ldots, c\}$ such that there are at most $k - l + |J| - 2$ nonzero elements among $\{g_\lambda^i : \lambda \in J,\ i = k, \ldots, n\}$. This implies that there are at most $k - l + |J| - 2$ nonzero columns of the matrix $M_P^J := \left(g_\lambda^i(P)\right)_{\lambda \in J}^{k \le i \le n}$. We can choose J so that the number $|J|$ is minimal among such subsets of $\{l+1, \ldots, c\}$. Then, by elementary operations as in the first part of the proof, we can transfer M_P^J to the form

$$\begin{pmatrix} h_{l+1}^k & \cdots & \cdots & h_{l+1}^{k+|J|} & \cdots & h_{l+1}^n \\ & \ddots & & \vdots & & \vdots \\ 0 & & h_{l+|J|}^{k+|J|-1} & h_{l+|J|}^{k+|J|} & \cdots & h_{l+|J|}^n \end{pmatrix}(P).$$

Note that on the bottom row we have at most $k - l - 1$ nonzero entries. Hence we obtain

$$\begin{aligned} \dim(f_1 = \cdots = f_l = 0) \cap (h_{l+|J|}^{k+|J|-1} = \cdots = h_{l+|J|}^n = 0) \cap \Pi^0 & \\ \ge k - l - (k - l - 1) = 1. & \end{aligned}$$

Since the rank of M_P^J is not maximal on the subset $(h_{l+|J|}^{k+|J|-1} = \cdots = h_{l+|J|}^n = 0)$, we see that C_X^* is singular along the above positive dimensional subset. Hence X is not quasismooth in this case. This concludes the proof of Proposition 3.1. □

In the following example, we use Proposition 3.1 to check quasismoothness of a given WCI.

Example 3.3. Let $X_{8,8,8} \subset \mathbb{P}(2^{(4)}, 3^{(5)}, 5^{(3)})$ be a general WCI of codimension 3. We can check the quasismoothness of $X_{8,8,8}$ by Proposition 3.1 as follows. Consider

$I = \{4, 5, 6, 7, 8\}$, that is, $a_4 = \cdots = a_8 = 3$. Then (Q1) does not hold for this I and we have $k = 5$, $l = 0$ in (Q2). We can choose $\{e_{\mu,j} : j = 1, 2, 3,\ \mu = 1, \ldots, 5\} \subset \{0, 1, 2, 3, 9, 10, 11\}$ so that (Q2) is satisfied for this I. We can similarly check that (Q2) holds for $I = \{9, 10, 11\}$. For other I, we have (Q1), thus we see the quasismoothness of $X_{8,8,8}$.

On the other hand, we see that $X'_{8,8,8} \subset \mathbb{P}(2^{(3)}, 3^{(4)}, 5^{(3)})$ is not quasismooth. Indeed, for $I = \{7, 8, 9\}$, that is, $a_7 = a_8 = a_9 = 5$, neither (Q1) nor (Q2) hold.

The following proposition treats the special situation where some weight of $\mathbb{P}$ divides none of the degrees of a WCI.

Proposition 3.4. *Let $X = X_{d_1,\ldots,d_c} \subset \mathbb{P}(a_0, \ldots, a_n)$ be a (well-formed) quasismooth WCI which is not a linear cone. Assume that there exists i_0 such that a_{i_0} does not divide d_j for all j. Let $H = \mathcal{O}_X(h)$ be the fundamental divisor on X, that is, an ample Cartier divisor on X which generates* Pic *X. Then*

(i) *X has a quotient singularity of type $1/a_{i_0}(c_1,\ldots,c_{n-c})$ for some $c_1,\ldots,c_{n-c} \in \mathbb{Z}_{\geq 0}$ such that* $\gcd(a_{i_0},c_1,\ldots,c_{n-c}) = 1$;

(ii) *$a_{i_0} \mid h$. As a consequence, we have $|H| \neq \emptyset$.*

Proof. Let $f_1, \ldots, f_c \in \mathbb{C}[x_0, \ldots, x_n]$ be the defining equations of X such that $\deg f_j = d_j$ for $1 \leq j \leq c$ and $X = (f_1 = \cdots = f_c = 0) \subset \mathbb{P}(a_0, \ldots, a_n)$, where $\deg x_i = a_i$ for $0 \leq i \leq n$. By applying Proposition 3.1 to $I = \{i_0\}$, we see that there exist distinct integers $e_1, \ldots, e_c \in \{0, \ldots, \hat{i_0}, \ldots, n\}$ and positive integers $k_1, \ldots, k_c$ such that $d_j = k_j a_{i_0} + a_{e_j}$ for $1 \leq j \leq c$, i.e., we can write

$$f_j = x_{i_0}^{k_j} x_{e_j} + g_j$$

for $1 \leq j \leq c$, where g_j is a weighted homogeneous polynomial of degree d_j.

By the inverse function theorem, we see that X has a quotient singularity of type $1/a_{i_0}(a_0, \ldots, \hat{a}_{i_0}, \ldots, \hat{a}_{e_1}, \ldots, \hat{a}_{e_c}, \ldots, a_n)$ at $P_{i_0} := [0 : \cdots : 1 : \cdots : 0]$. We shall show that $g := \gcd(a_0, \ldots, \hat{a}_{e_1}, \ldots, \hat{a}_{e_c}, \ldots, a_n) = 1$. Suppose that $g > 1$.

Claim 3.5. *Up to a permutation on $\{1, \ldots, c\}$, we may choose $0 \leq c' \leq c$ with the following properties*:

(*) *For $j = 1, \ldots, c'$, some monomial in g_j does not contain any element of $\{x_{e_j}, \ldots, x_{e_c}\}$.*

(**) *For $j = c' + 1, \ldots, c$, every monomial in g_j contain some of $\{x_{e_{c'+1}}, \ldots, x_{e_c}\}$.*

Proof of the claim. If (**) holds for all $j = 1, \ldots, c$ and $\{x_{e_1}, \ldots, x_{e_c}\}$, then we put $c' := 0$. Otherwise there is some j such that $1 \leq j \leq c$ and (*) holds for $\{x_{e_1}, \ldots, x_{e_c}\}$. We then exchange (f_1, e_1) and (f_j, e_j) and repeat the same process starting from $j = 2$ till we obtain the claim, that is, check whether (**) holds for new $\{f_2, \ldots, f_c\}$ and $\{e_2, \ldots, e_c\}$ and so on. $\square$

Hence, for $1 \le j \le c'$, there exists a monomial in g_j of the form $h_j = \prod_{i \neq e_j, \dots, e_c} x_i^{b_i}$. Then we have $a_{e_j} \equiv \sum_{i \neq e_j, \dots, e_c} b_i a_i \mod a_{i_0}$. Thus we can check one by one that

$$g \mid a_{e_j} \quad \text{for } 1 \le j \le c'. \tag{3}$$

Now let $\Pi := (x_{e_{c'+1}} = \cdots = x_{e_c} = 0) \subset \mathbb{P}$. We have $\Pi \subset \operatorname{Sing} \mathbb{P}$, in particular $\Pi \neq \mathbb{P}$.

We also have $f_j|_\Pi \equiv 0$ for $c'+1 \le j \le c$ by the property (**) of c'. Thus we obtain

$$\dim \Pi \cap X \ge \dim \Pi - c' = \dim \mathbb{P} - c.$$

This contradicts the fact that $X \not\subset \Pi$ since X is not a linear cone. Hence we obtain $g = 1$, concluding the proof of Proposition 3.4. □

The following proposition is useful for calculating the fundamental divisor of a WCI and is the motivation of the definition of *h-regular pair* (see Definition 4.1).

Proposition 3.6. *Let $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a quasismooth well-formed WCI which is not a linear cone. Let $H = \mathcal{O}_X(h)$ be the fundamental divisor of X. Assume that there exists $I = \{i_1, \dots, i_k\}$ such that $a_I := \gcd(a_{i_1}, \dots, a_{i_k}) > 1$.*

Then one of the following holds:

(i) *There exist distinct integers $p_1, \dots, p_k$ such that $a_I \mid d_{p_1}, \dots, d_{p_k}$;*

(ii) *$a_I \mid h$.*

Proof. We apply Proposition 3.1 to $I = \{i_1, \dots, i_k\}$. Let

$$P_I := (x_0 = \cdots \hat{x}_{i_1} = \cdots = \hat{x}_{i_k} = \cdots = x_n = 0) \subset \mathbb{P}$$

be the $(k-1)$-dimensional stratum corresponding to I and $P_I^0 \subset P_I$ be the open toric stratum.

Suppose that condition (Q1) in Proposition 3.1 holds, that is, there exist distinct integers $p_1, \dots, p_k$ and nonnegative integers $k_{j,i}$ for $j = 1, \dots, k$ and $i \in I$ such that $d_{p_j} = \sum_{i \in I} k_{j,i} a_i$. Then we have (i) in this case.

Suppose that (Q2) holds. Then there exist a permutation $p_1, \dots, p_c$ of $\{1, \dots, c\}$, an integer $l < \rho := \min\{c, k\}$, nonnegative integers $k_{j,i}$ for $j = 1, \dots, c$ and $i \in I$, and distinct integers $e_{l+1}, \dots, e_c$, which satisfy the following:

- for $j = 1, \dots, l$, we have $\sum_{i \in I} k_{j,i} a_i = d_{p_j}$,
- for $j = l+1, \dots, c$, we have $a_{e_j} + \sum_{i \in I} k_{j,i} a_i = d_{p_j}$.

We may assume that $(f_{p_j} = 0) \cap P_I^0 \neq \varnothing$ since X is irreducible and the linear system $|\mathcal{O}_\mathbb{P}(d_{p_j})|$ does not have a fixed component. Hence, on $p \in X \cap P_I^0$, the variety X is analytic locally isomorphic to a quotient singularity of type

$$\frac{1}{a_I}(a_0, \dots, \hat{a}_{i_1}, \dots, \hat{a}_{i_k}, \dots, \hat{a}_{e_{l+1}}, \dots, \hat{a}_{e_c}, \dots, a_n) \times \mathbb{C}^{k-l}.$$

Now the proof is reduced to the following claim:

Claim 3.7. *We have* $g := \gcd(a_0, \dots, \hat{a}_{e_{l+1}}, \dots, \hat{a}_{e_c}, \dots, a_n) = 1$.

Proof of Claim. Suppose that $g > 1$. We shall have a similar contradiction as in the proof of Proposition 3.4. As in Claim 3.5, up to a permutation of $\{1, \dots, c\}$, we may choose c' with $l+1 \le c' \le c$ with the following properties:

(*) For $j = l+1, \dots, c'$, some monomial in g_j does not contain any element of $\{x_{e_j}, \dots, x_{e_c}\}$.

(**) For $j = c'+1, \dots, c$, every monomial in g_j contain some of $\{x_{e_{c'+1}}, \dots, x_{e_c}\}$.

Let $\Pi := (x_{e_{c'+1}} = \dots = x_{e_c} = 0) \subset \mathbb{P}$. Then, as in Proposition 3.6, we see that $f_j|_\Pi \equiv 0$ for $j = c'+1, \dots, c$ and $\Pi \subset \operatorname{Sing} \mathbb{P}(a_0, \dots, a_n)$ since $g \mid a_i$ for $i \notin \{e_{c'+1}, \dots, e_c\}$. Thus we have $\dim \Pi \cap X \ge \dim \mathbb{P} - c$ as before and it contradicts that $X \not\subset \Pi$ since X is not a linear cone. Thus we obtain the claim. □

The sheaf $\mathcal{O}_X(1)$ induces a generator of the class group of a quotient singularity of the above type. Since the class group is a cyclic group of order a_I as in (1), we see that $a_I \mid h$. Thus we have finished the proof of Proposition 3.6. □

The following corollary restricts Proposition 3.6 to the smooth case.

Corollary 3.8 [Przyjalkowski and Shramov 2016, Lemma 2.15]. *Let* $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ *be a smooth WCI. Assume that there exists* $I = \{i_1, \dots, i_k\}$ *such that* $a_I := \gcd(a_{i_1}, \dots, a_{i_k}) > 1$.

Then there exist distinct integers $p_1, \dots, p_k$ *such that* $a_I \mid d_{p_1}, \dots, d_{p_k}$.

Proof. Since X is smooth, the fundamental divisor of X is $\mathcal{O}_X(1)$, that is $h = 1$ in the notation of Proposition 3.6. Thus the statement follows from Proposition 3.6. □

4. Regular pairs and Frobenius coin problem

The following definition is motivated by Proposition 3.6 and Corollary 3.8.

Definition 4.1. Let $c \in \mathbb{N}$ and $n \in \mathbb{Z}_{\ge -1}$ be integers and $(d; a)$ be a pair, where $d = (d_1, \dots, d_c) \in \mathbb{N}_+^c$ and $a = (a_0, \dots, a_n) \in \mathbb{N}_+^{n+1}$. Let $\bar{c}^+ := \{1, \dots, c\}$ and $\bar{n} := \{0, \dots, n\}$.

We say that $(d; a)$ is *h-regular* for a positive integer h if, for any subset $I = \{i_1, \dots, i_k\} \subset \bar{n}$ such that $a_I := \gcd(a_{i_1}, \dots, a_{i_k}) > 1$, one of the following holds:

(i) There exist distinct integers $p_1, \dots, p_k \in \bar{c}^+$ such that $a_I \mid d_{p_1}, \dots, d_{p_k}$;

(ii) $a_I \mid h$.

If a pair is h-regular for $h = 1$, we simply call it *regular*.

Remark 4.2. For technical reasons, in Definition 4.1 we admit the cases $c = 0$ or $n = -1$, i.e., pairs of the form $(d; \varnothing)$, $(\varnothing; a)$ and $(\varnothing, \varnothing)$.

We need to fix some notation. If $(d; a)$ is a pair with $d = (d_1, \dots, d_c) \in \mathbb{N}_+^c$ and $a = (a_0, \dots, a_n) \in \mathbb{N}_+^{n+1}$, then we define

$$\delta(d; a) := \sum_{j=1}^{c} d_j - \sum_{i=0}^{n} a_i.$$

In the case where the pair $(d; a)$ comes from a well-formed quasismooth WCI $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$, we have $\omega_X \cong \mathcal{O}_X(\delta(d; a))$.

Let q be a prime number. Set $I_q := \{i \in \bar{n} : q \mid a_i\}$ and $J_q := \{j \in \bar{c}^+ : q \mid d_j\}$. We consider two new pairs obtained from $(d; a)$. The pair $(d^q; a^q)$ is given by

$$d^q := ((d_j/q)_{j\in J_q}, (d_j)_{j\in\bar{c}^+\setminus J_q}), \quad a^q := ((a_i/q)_{i\in I_q}, (a_i)_{i\in\bar{n}\setminus I_q})$$

in which we divided by q all the divisible d_j and a_i and the pair $(d(q), a(q))$ is given by

$$d(q) := (d_j)_{j\in J_q}, \quad a(q) := (a_i)_{i\in I_q}$$

in which only the divisible d_j and a_i appear. Note that

$$\delta(d; a) = \delta(d^q; a^q) + \frac{q-1}{q}\delta(d(q); a(q)).$$

Definition 4.3. For a pair $(d; a)$, we may choose subsets $J_{(d;a)} = \{j_1,\dots, j_l\} \subset \bar{c}^+$ and $I_{(d;a)} = \{i_1,\dots,i_l\} \subset \bar{n}$ uniquely for some $l \in \mathbb{N}$ so that $d_{j_k} = a_{i_k}$ for $k = 1,\dots,l$ and $d_j \neq a_i$ for all $j \in \bar{c}^+ \setminus J_{(d;a)}$ and $i \in \bar{n} \setminus I_{(d;a)}$. We define a pair $(\tilde{d}; \tilde{a})$ by

$$(\tilde{d}; \tilde{a}) := ((d_j)_{j\in\bar{c}^+\setminus J_{(d;a)}}; (a_i)_{i\in\bar{n}\setminus I_{(d;a)}}), \tag{4}$$

that is, we cancel the doubles (d_j, a_i) with $d_j = a_i$.

Lemma 4.4. *The pair $(\tilde{d}; \tilde{a})$ is h-regular if $(d; a)$ is h-regular.*

Proof. Let $I := \{i_1, \dots, i_k\} \subset \bar{n} \setminus I_{(d;a)}$ be a subset with $a_I > 1$. Since $(d; a)$ is h-regular, either (i) holds for some $\{p_1, \dots, p_k\} \subset \bar{c}^+$ or (ii) holds. In the latter case, there is nothing to check. Thus we consider the former case and need to find $p'_1, \dots, p'_k \in \bar{c}^+ \setminus J_{(d;a)}$ such that $a_I \mid d_{p'_j}$ for $j = 1, \dots, k$. Let

$$J' := \{j \in J_{(d;a)} : a_I \mid d_j\}, \quad I' := \{i \in I_{(d;a)} : a_I \mid a_i\}.$$

Then we have $|I'| = |J'| =: l'$. Let $I'' := I \cup I'$. By $a_I = a_{I''}$, there exist distinct integers $p_1, \dots, p_{k+l'} \in \bar{c}^+$ such that $a_I \mid d_{p_j}$ for $j = 1, \dots, k + l'$. Then the set $\{p_1, \dots, p_{k+l'}\} \setminus J'$ contains k elements $p'_1, \dots, p'_k \in \bar{c}^+ \setminus J_{(d;a)}$ such that $a_I \mid d_{p'_j}$ for $j = 1, \dots, k$. Thus (i) holds for $(\tilde{d}; \tilde{a})$ and I. Hence we see that $(\tilde{d}; \tilde{a})$ is h-regular. $\square$

The following straightforward lemmas show how h-regular pairs are very suitable for inductive arguments.

Lemma 4.5. *Let $(d; a)$ be an h-regular pair and q be a prime not dividing h. Then the pairs $(d^q; a^q)$ and $(d(q); a(q))$ are h-regular. Hence $(d(q)/q; a(q)/q)$ is also h-regular.*

Proof. We write the details for the pair $(d^q; a^q)$. The proof for $(d(q); a(q))$ is easier.

Let $I = \{i_1, \dots, i_k\} \subset \bar{n}$ such that $a_I := \gcd(a_{i_1}, \dots, a_{i_k}) > 1$. By the h-regularity of $(d; a)$, we have either condition (i) of Definition 4.1, i.e., there exist distinct integers $p_1, \dots, p_k$ such that

$$a_I \mid d_{p_1}, \dots, d_{p_k}$$

or (ii), i.e., $a_I \mid h$.

If $q \mid a_I$, then we have $q \mid a_{i_\ell}$ for all $i_\ell \in I$ and $a_I \nmid h$. Thus we have (i) and

$$\gcd\Big(\frac{a_{i_1}}{q}, \dots, \frac{a_{i_k}}{q}\Big) = \frac{a_I}{q} \;\Big|\; \frac{d_{p_1}}{q}, \dots, \frac{d_{p_k}}{q}$$

as we wanted.

If $q \nmid a_I$, then

$$\gcd((a_i/q)_{i \in I'_q}, (a_i)_{i \in (I \setminus I'_q)}) = a_I$$

where $I'_q := \{i \in I : q \mid a_i\}$. If $a_I \mid h$, then there is nothing to prove, so we can assume that $a_I \nmid h$ and that (i) holds. Since $q \nmid a_I$, we get that a_I divides $(d^q)_{p_j}$ for $j = 1, \dots, k$. This concludes the proof. □

Lemma 4.6. *Let $(d; a)$ be an h-regular pair and q be a prime dividing h. Then $(d^q; a^q)$ is h/q-regular and $(d(q); a(q))$ is h-regular, hence $(d(q)/q; a(q)/q)$ is h/q-regular.*

Proof. We give the proof for the pair $(d^q; a^q)$. Consider a set $I = \{i_1, \dots, i_k\} \subset \bar{n}$ such that $a_I := \gcd(a_{i_1}, \dots, a_{i_k}) > 1$ and let $a_I^q := \gcd(a_i^q)_{i \in I}$ be the gcd of the a_i in $(d^q; a^q)$.

Assume first that $\gcd(a_I, q) = 1$, so that $a_I^q = a_I$. If $a_I \mid h$, we obtain that $a_I^q \mid h/q$ and we are done. If $a_I \nmid h$, then there exist distinct integers $p_1, \dots, p_k$ such that $a_I \mid d_{p_1}, \dots, d_{p_k}$. We have $a_I^q \nmid h/q$ and the $d_{p_j}^q$ work.

If $a_I = qt$ for some positive integer t, then $a_I^q = t$. If $qt \mid h$, we have $t \mid h/q$. If $qt \nmid h$, then there exist distinct integers $p_1, \dots, p_k$ such that $a_I \mid d_{p_1}, \dots, d_{p_k}$. For the same integers, we have $a_I^q \mid d_{p_1}/q, \dots, d_{p_k}/q$ so the first condition of h/q-regularity is satisfied and we are done. □

The Frobenius coin problem. In this subsection we want to enlighten some interesting connections among the Ambro–Kawamata conjecture, regular pairs and the Frobenius coin problem.

Question 4.7 (Frobenius coin problem). Given positive integers $a_0, \ldots, a_n$ such that $\gcd(a_0, \ldots, a_n) = 1$, find the largest integer $G = G(a_0, \ldots, a_n)$ so that there do not exist nonnegative integers $x_0, \ldots, x_n$ satisfying

$$G = a_0 x_0 + \ldots + a_n x_n.$$

Such G is called the Frobenius number of $a_0, \ldots, a_n$.

For $n = 1$, it is classically known that

$$G(a_0, a_1) = a_0 a_1 - a_0 - a_1.$$

For $n \geq 2$, the problem is considerably harder: precise methods have been developed to compute $G(a_0, a_1, a_2)$ and some algorithms and (lower and upper) bounds are known for the general case (see for instance [Johnson 1960] and [Brauer and Shockley 1962]).

By Lemma 2.4, the Ambro–Kawamata conjecture for smooth WCI would follow from the following purely arithmetic statement, which we believe to be of independent interest.

Conjecture 4.8. *Let $(d; a) = (d_1, \ldots, d_c; a_0, \ldots, a_n) \in \mathbb{N}^c \times \mathbb{N}^{n+1}$ be a regular pair such that $a_i \neq 1$ and $d_j \neq a_i$ for any i, j. Assume $c \leq n$ and $\gcd(a_0, \ldots, a_n) = 1$. Then*

$$\delta(d; a) \geq G(a_0, \ldots, a_n).$$

One of the best known lower bounds for G is given in [Brauer 1942]. Let $a_0, \ldots, a_n$ be positive coprime integers, set $g_j := \gcd(a_0, \ldots, a_j)$ for $j = 0, \ldots, n$ and consider

$$Br(a_0, \ldots, a_n) := \sum_{j=1}^{n} a_j \frac{g_{j-1}}{g_j} - \sum_{i=0}^{n} a_i.$$

Brauer proved that $Br(a_0, \ldots, a_n) \geq G(a_0, \ldots, a_n)$. Set $d_j := a_j g_{j-1}/g_j$ for $j = 1, \ldots, n$. Then it is easy to check that $(d; a) := (d_1, \ldots, d_n; a_0, \ldots, a_n)$ is actually a regular pair.

On the other hand, it is not difficult to see that, considering big prime numbers p and q, the pair $(pq, 6p, 6q; 2p, 3p, 2q, 3q)$ is regular, $\delta(d; a) \geq G(a_0, \ldots, a_n)$, but $\delta(d; a) < Br(a_0, \ldots, a_n)$.

This shows that regular pairs can give better bounds for the Frobenius number with respect to the known ones. For this reason, it seems to be a challenge and interesting problem to study Conjecture 4.8.

Remark 4.9. It is not difficult to check that Conjecture 4.8 is true for $c = 1, 2$, which implies that the nonvanishing holds for a smooth WCI of codimension 1 or 2. For simplicity, we omit the detail in the codimension 2 case.

For $c=1$, a stronger and more general result is given in Lemma 6.1, which is the key step to prove Theorem 1.3.

5. Proof of Theorem 1.2

Theorem 1.2 is the combination of Corollary 5.3 and Corollary 5.13 below.

Smooth case. The pair $(d;a)$ in the following lemma does not come from a nonempty WCI. Nevertheless this lemma is important in the proof of Proposition 5.2.

Lemma 5.1. *Let $(d;a) \in \mathbb{N}_+^c \times \mathbb{N}_+^{n+1}$ be a regular pair such that $a_i \neq d_j$ for any i, j. Let q be a prime number such that $q \mid a_i$ and $q \mid d_j$ for any i, j. Then*

$$\delta(d;a) \geq cq.$$

Moreover, if the equality holds, then $c=n+1$.

Proof. Note that $c \geq n+1$ which does not occur for a nonempty WCI.

Assume first that q is the only prime dividing the a_i, that is for any $i=0,\dots,n$, we have $a_i = q^{\alpha_i}$ for some $\alpha_i \geq 1$. We can assume that $a_0 \geq a_1 \geq \dots \geq a_n$. We can also order the d_j in such a way that $v_q(d_s) \geq v_q(d_t)$ for any $s \leq t$, where $v_q(d_j) = \max\{e \in \mathbb{N} : q^e \mid d_j\}$. Then we have $a_i \mid d_{i+1}$ for any $i=0,\dots,n$ and so

$$\sum_{j=1}^{c} d_j - \sum_{i=0}^{n} a_i = \sum_{k=1}^{c-n-1} d_{n+1+k} + \sum_{i=0}^{n} (d_{i+1} - a_i) \geq cq$$

and the equality is possible only if $c=n+1$, $d_j = 2q$ and $a_i = q$ for any i, j.

Assume now that $q \neq 2$ and that q and 2 are the only primes dividing the a_i, that is for any $i=0,\dots,n$ we have $a_i = 2^{\alpha_i} q^{\beta_i}$ for some $\alpha_i \geq 0$ and $\beta_i \geq 1$ such that $\alpha_i > 0$ for at least one i. We proceed by induction on $t = \max_{0 \leq i \leq n}\{\beta_i\}$, the greatest power of q dividing at least one a_i.

Suppose $t=1$. We can assume that $v_2(a_i) \geq v_2(a_j)$ and $v_2(d_i) \geq v_2(d_j)$ for any $i \leq j$. Then again $a_i \mid d_{i+1}$ for any $i=0,\dots,n$ and we conclude as before.

Suppose $t \geq 2$. Let $I_{q^t} := \{i \in \bar{n} : q^t \mid a_i\}$ and $J_{q^t} := \{j \in \bar{c}^+ : q^t \mid d_j\}$. We consider the following pairs: $(d';a')$, where $d' = ((d_j/q)_{j \in J_{q^t}}, (d_j)_{j \in \bar{c}^+ \setminus J_{q^t}})$ and $a' = ((a_i/q)_{i \in I_{q^t}}, (a_i)_{i \in \bar{n} \setminus I_{q^t}}))$ and $(d'';a'')$, where $d'' = (d_j/q)_{j \in J_{q^t}}$ and $a'' = (a_i/q)_{i \in I_{q^t}}$. It is straightforward to check as in Lemma 4.5 that $(d';a')$ and $(d'';a'')$ are regular. Consider the regular pair $(\tilde{d}';\tilde{a}')$ constructed in (4) which satisfies $\tilde{d}'_j \neq \tilde{a}'_i$ for any $i \in \bar{n} \setminus I_{(d';a')}$, $j \in \bar{c}^+ \setminus J_{(d';a')}$, where $I_{(d';a')} \subset \bar{n}$ and $J_{(d';a')} \subset \bar{c}^+$ are the subsets defined in Definition 4.3.

Let

$$m := |\{j \in J_{q^t} : d_j/q = a_i \mid \text{for some } i \in \bar{n} \setminus I_{q^t}\}|,$$

$$\bar{m} := |\{i \in I_{q^t} : d_j = a_i/q \mid \text{for some } j \in \bar{c}^+ \setminus J_{q^t}\}|.$$

Note that $|I_{(d';a')}| = |J_{(d';a')}| \leq m+\overline{m}$. Let $k := |J_{q^t}|$. By induction on t, we may assume that we have $\delta(d';a') = \delta(\tilde{d}';\tilde{a}') \geq (c-m-\overline{m})q$ and $\delta(d'';a'') \geq kq$. We have that $|I_{q^t}| \leq |J_{q^t}|$ since $(d;a)$ is regular. Since $m \leq k$ and $\overline{m} \leq |I_{q^t}| \leq k$, we obtain

$$\begin{aligned}\delta(d;a) = \delta(d';a') + (q-1)\delta(d'';a'') &\geq (c-2k)q + (q-1)kq \\ &= cq - 2kq + kq^2 - kq \\ &= cq + kq(q-3) \geq cq \qquad (5)\end{aligned}$$

because $q \geq 3$. The equality is possible only if we have it for both $(\tilde{d}';\tilde{a}')$ and $(d'';a'')$. This implies by induction on t that $c = n+1$ in this case.

We now pass to the general case. For any prime p, different from q and 2, let $e_p := \max\{e \in \mathbb{N} : p^e \mid a_i \text{ for some } i\}$. The proof is by induction on $D = \sum_p e_p$, where the index varies over all prime numbers different from q and 2. The case $D = 0$ has already been treated in the first part of the proof. So assume $D \geq 1$ and that the inequality holds up to $D-1$. Consider $(d^p;a^p)$ and let

$$\begin{aligned}m_p &:= |\{j \in \bar{c}^+ : d_j/p = a_i \text{ for some } i \in \bar{n} \setminus I_p\}|,\\ \overline{m}_p &:= |\{i \in \bar{n} : d_j = a_i/p \text{ for some } j \in \bar{c}^+ \setminus J_p\}|.\end{aligned}$$

Let us again consider the pair $(\tilde{d}^p;\tilde{a}^p)$ as in Definition 4.3 by removing subsets $J_{(d^p;a^p)} \subset \bar{c}^+$ and $I_{(d^p;a^p)} \subset \bar{n}$. Then this satisfies the hypothesis $(\tilde{d}^p)_j \neq (\tilde{a}^p)_i$ for any i and j. We again have that $|J_{(d^p;a^p)}| \leq m_p + \overline{m}_p$. By induction on D, we obtain $\delta(d^p;a^p) = \delta(\tilde{d}^p;\tilde{a}^p) \geq (c-m_p-\overline{m}_p)q$. Now consider the pair $(d(p)/p;a(p)/p)$. Again by induction on D, we obtain $\delta(d(p)/p;a(p)/p) \geq sq$, where $s := |\{j \in \bar{c}^+ : p|d_j\}|$. We see that $m_p \leq s$ by the definition of m_p. Let $s' := |\{i \in \bar{n} : p \mid a_i\}|$. We see that $s' \leq s$ by the regularity of $(d;a)$ and that $\overline{m}_p \leq s'$ by the definition of $\overline{m}_p$. Thus we have $\overline{m}_p \leq s$. By these inequalities and $p \geq 3$, we conclude that

$$\begin{aligned}\delta(d;a) &= \delta(d^p;a^p) + (p-1)\delta(d(p)/p;a(p)/p)\\ &\geq (c-m_p-\overline{m}_p)q + (p-1)sq\\ &= cq + psq - m_pq - \overline{m}_pq - sq \geq cq + psq - 3sq \geq cq\end{aligned}$$

as we wanted. Again, the equality is possible only if $c = n+1$. □

By using Lemma 5.1, we prove the following key proposition.

Proposition 5.2. *Let $(d;a) \in \mathbb{N}_+^c \times \mathbb{N}_+^{n+1}$ be a regular pair such that $a_i > 1$ and $a_i \neq d_j$ for any i, j. Then the following holds.*

(i) *We have*

$$\delta(d;a) \geq c. \qquad (6)$$

(ii) *If $\gcd(a_0, \dots, a_n) = 1$, then the equality holds only if $(d;a)$ is of the form $(6^{(s)}, 1^{(c-s)}; 2^{(s)}, 3^{(s)})$ for some integer s.*

Proof. (i) The proof is by induction on n, the case $n = 0$ being obvious. We can assume that no prime divides every a_i, otherwise we are in the case of Lemma 5.1. In particular, we may assume that there is a prime $q \neq 2$ which divides some a_i. Let

$$m := |\{j \in \bar{c}^+ : d_j/q = a_i \text{ for some } i \in \bar{n} \setminus I_q\}|,$$
$$\overline{m} := |\{i \in \bar{n} : d_j = a_i/q \text{ for some } j \in \bar{c}^+ \setminus J_q, d_j \neq 1\}|,$$
$$\ell := |\{i \in \bar{n} : a_i = q\}|, \quad s := |\{j \in \bar{c}^+ : q | d_j\}| = |J_q|.$$

We note that $m \leq s$ by definition and $\ell + \overline{m} \leq s$ by the regularity.

Case 1: Suppose that $\ell + m + \overline{m} \geq 1$. Then the pair $(d^q; a^q)$ has some redundant a_i, in the sense that $a_i/q = 1$, $d_j/q = a_i$ or $d_j = a_i/q$ for some i, j. That is, we consider a regular pair $(\tilde{d}^q, \tilde{a}^q)$ and, by removing all $\tilde{a}^q_i = 1$, we obtain a new regular pair $(\hat{d}^q; \hat{a}^q) \in \mathbb{N}_+^{\hat{c}} \times \mathbb{N}_+^{\hat{n}+1}$ for some $\hat{c} \leq c$ and $\hat{n} \leq n$. Note that $\hat{n} < n$ by the hypothesis $\ell + m + \overline{m} \geq 1$. Let $\ell_1 := |\{j \in \bar{c}^+ : d_j = 1\}|$ and $\ell' := \min\{\ell, \ell_1\}$. Then we see that $\hat{c} \geq c - m - \overline{m} - \ell'$ by the construction of $(\tilde{d}^q; \tilde{a}^q)$. Since we have $|\{i \in \bar{n} \setminus I_{(d^q;a^q)} : \tilde{a}^q_i = 1\}| = \ell - \ell'$, we obtain, by induction on n, that

$$\delta(d^q; a^q) = \delta(\hat{d}^q; \hat{a}^q) - (\ell - \ell') \geq \hat{c} - (\ell - \ell') \geq c - \ell - m - \overline{m}.$$

By applying Lemma 5.1 to $(d(q), a(q))$, we obtain

$$\delta(d(q); a(q)) \geq sq.$$

By these and $\ell + m + \overline{m} \leq 2s$, we obtain

$$\begin{aligned} \delta(d; a) = \delta(d^q; a^q) + \frac{q-1}{q}\delta(d(q); a(q)) &\geq c - \ell - m - \overline{m} + (q-1)s \\ &\geq c + qs - 3s \geq c \end{aligned} \tag{7}$$

since $q \geq 3$.

Case 2: Suppose now that $\ell + m + \overline{m} = 0$. Then the pair $(d^q; a^q)$ satisfies the assumptions of the proposition. We note that

$$\delta(d; a) = \delta(d^q; a^q) + \frac{q-1}{q}\delta(d(q); a(q)) > \delta(d^q; a^q)$$

since we have $\delta(d(q); a(q)) > 0$ by Lemma 5.1. So we can replace the pair $(d; a)$ with $(d^q; a^q)$ without changing the number c of the j and we can repeat the argument from the beginning of the proof (possibly changing the prime q) till either we end up in **Case 1** or we reach the situation of Lemma 5.1. In both cases, we are done and obtain (6).

(ii) We now study when the identity holds in the case $\gcd(a_0, \ldots, a_n) = 1$.

Note that the case $n = 1$ is clear, being equivalent to asking $a_0 a_1 - a_0 - a_1 = 1$.

Assume $n \geq 2$ and let $q \neq 2$ be a prime number such that $q \mid a_i$ for some i. We shall follow the proof of the inequality. In particular, we look at

$$\delta(d; a) = \delta(d^q; a^q) + \frac{q-1}{q}\delta(d(q); a(q)).$$

With the same notation as above, we note that the equality can hold only if we are in **Case 1** and, by Lemma 5.1, the number $|\{i \in \bar{n} : q \mid a_i\}|$ must be equal to $s = |J_q|$. Moreover we obtain $q = 3$ by (7). This implies that the only possible prime numbers that divide at least one a_i are 2 and 3.

We must also have $m = s$ and $\ell + \overline{m} = s$. By $m = s$, we see that any $d_j/3 \in \mathbb{N}$ must be equal to some a_i which is not divisible by 3. Hence we can write

$$(d; a) = (3 \cdot 2^{\beta_1}, \dots, 3 \cdot 2^{\beta_s}, 2^{\beta_{s+1}}, \dots, 2^{\beta_c}; 2^{\beta_1}, \dots, 2^{\beta_s}, 3 \cdot 2^{\alpha_{s+1}}, \dots, 3 \cdot 2^{\alpha_{n+1}})$$

for some nonnegative integers α_i and β_i. Then

$$\delta(d;a) = \delta(2^{\beta_{s+1}}, \dots, 2^{\beta_c}; 2^{\alpha_{s+1}}, \dots, 2^{\alpha_{n+1}}) + \tfrac{2}{3}\delta(3 \cdot 2^{\beta_1}, \dots, 3 \cdot 2^{\beta_s}; 3 \cdot 2^{\alpha_{s+1}}, \dots, 3 \cdot 2^{\alpha_{n+1}}).$$

Also note that $\ell + \overline{m} = s$ implies that $n + 1 - s = |\{i \in \bar{n} : 3 \mid a_i\}| = \ell + \overline{m} = s$, thus $n + 1 = 2s$. By the regularity of $(d(3); a(3))$ and the assumption $d_j \neq a_i$ for any i, j, to have the equality $\delta(d(3); a(3)) = 3s$ we need $\beta_j = 1$ for $j = 1, \dots, s$ and $\alpha_i = 0$ for $i = s + 1, \dots, n + 1$, which implies

$$\delta(d; a) = \delta(2^{\beta_{s+1}}, \dots, 2^{\beta_c}; 1, \dots, 1) + 2s = \sum_{i=1}^{c-s} 2^{\beta_{s+i}} - (n + 1 - s) + 2s,$$

i.e., $c = \delta(d; a) = \sum_{i=1}^{c-s} 2^{\beta_{s+i}} + s$.

Hence, we must have $\beta_j = 0$ for $j = s + 1, \dots, c$, which finishes the proof. □

As a corollary of Proposition 5.2, we obtain the nonemptiness of $|\mathcal{O}(1)|$ and the smoothness of its general member on a smooth Fano or Calabi–Yau WCI.

Corollary 5.3. *Let $X := X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a well-formed smooth Fano or Calabi–Yau WCI which is not a linear cone. Let $c_1 := |\{i \in \bar{n} : a_i = 1\}|$. Then the following hold.*

(i) *We have $c_1 \geq c$. Moreover the equality is possible only if X is Calabi–Yau of type $X_{6,\dots,6} \subset \mathbb{P}(1^{(c)}, 2^{(c)}, 3^{(c)})$.*

(ii) *The linear system $|\mathcal{O}_X(1)|$ is nonempty and its general member H is smooth.*

Proof. (i) We may assume that $a_0 \leq \dots \leq a_n$. Thus we have $a_0 = \dots = a_{c_1-1} = 1$. Since X is smooth, we see that $(d_1, \dots, d_c; a_{c_1}, \dots, a_n)$ is regular. By this and Proposition 5.2(i), we obtain

$$\delta(d_1, \dots, d_c; a_{c_1}, \dots, a_n) \geq c.$$

By the assumptions, $0 \geq \delta(d; a) \geq c - c_1$, and this implies the former statement.

Let $(d; a)$ be a regular pair which satisfies $c_1 = c$. Let $(\hat{d}; \hat{a})$ be the regular pair obtained by removing all $a_i = 1$. Then $(\hat{d}; \hat{a})$ satisfies the hypothesis of Proposition 5.2(ii) since $(d; a)$ defines a smooth WCI. Hence, by Proposition 5.2(ii), we see that $(\hat{d}; \hat{a}) = (6^{(c)}; 2^{(c)}, 3^{(c)})$ and $(d; a) = (6^{(c)}; 2^{(c)}, 3^{(c)}, 1^{(c)})$.

(ii) By the latter part of Proposition 5.2, we can assume that X is not of the form $X_{6,\ldots,6} \subset \mathbb{P}(2^{(c)}, 3^{(c)}, 1^{(c)})$; otherwise the conclusion is immediate. In particular, we may assume $c_1 \geq c + 1$.

By (i), we see that $|\mathcal{O}_X(1)| \neq \varnothing$. Since X is smooth and well-formed, we have $\operatorname{Sing} \mathbb{P}(a_0, \ldots, a_n) \cap X = \varnothing$. Thus $H \cap \operatorname{Sing} \mathbb{P}(a_0, \ldots, a_n) = \varnothing$. Hence it is enough to check H is quasismooth at $P := \Pi(p)$, where $p \in \Pi^{-1}((x_0 = \ldots = x_{c_1-1} = 0) \cap X)$ and $\Pi : \mathbb{A}^{n+1} \setminus \{0\} \to \mathbb{P}(a_0, \ldots, a_n)$ is the quotient map.

Set $H_i := X \cap (x_i = 0)$ for $i = 0, \ldots, c_1 - 1$. We shall look at the Jacobi matrices of X and $H_i \subset \mathbb{P}(a_0, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. Let $f_1, \ldots, f_c$ be the defining equations of X such that $\deg f_j = d_j$. For $i = 0, \ldots, n$, set

$$\boldsymbol{v}_i(p) := \begin{pmatrix} \partial f_1/\partial x_i \\ \vdots \\ \partial f_c/\partial x_i \end{pmatrix} (p).$$

The Jacobi matrix $J_X(p)$ and $J_{H_i}(p)$ of X and H_i can be written as

$$J_X(p) = (\boldsymbol{v}_0(p), \ldots, \boldsymbol{v}_n(p)), \quad J_{H_i}(p) = (\boldsymbol{v}_0(p), \ldots, \boldsymbol{v}_{i-1}(p), \boldsymbol{v}_{i+1}(p), \ldots, \boldsymbol{v}_n(p)).$$

Since X is quasismooth, there exist linearly independent vectors

$$\boldsymbol{v}_{i_1}(p), \ldots, \boldsymbol{v}_{i_c}(p).$$

Since $c_1 \geq c + 1$, we can choose i so that $i \notin \{i_1, \ldots, i_c\}$. Then we see that H_i is quasismooth at $P := \Pi(p)$. Thus a general member H is also quasismooth at P. □

Remark 5.4. Let $X \subset \mathbb{P}(a_0, \ldots, a_n)$ be a smooth WCI as in Corollary 5.3. For $I \subset \bar{n}$ such that $a_I = 1$, it may a priori happen that (Q1) does not hold, but (Q2) holds. That is why we make an argument as in Corollary 5.3 (ii).

Remark 5.5. Let $X_{d_1,\ldots,d_c}$ be a smooth WCI as in Corollary 5.3. Motivated by a question by Andreas Höring, we consider the description of the base locus $\operatorname{Bs} |\mathcal{O}_X(1)|$.

Up to reordering $d_1, \ldots, d_c$, we can assume that there is an integer $c' \leq c$ with the following properties: for $1 \leq j \leq c'$, there are weighted homogeneous polynomials $f_j(x_{c_1}, \ldots, x_n)$ of degree d_j and, for $c' + 1 \leq j \leq c$, all monomials of degree d_j contain one of the variables $x_0, \ldots, x_{c_1-1}$ of weights 1. Since the base locus $\operatorname{Bs} |\mathcal{O}_X(1)|$ is $(x_0 = \cdots = x_{c_1-1} = 0) \cap X_{d_1,\ldots,d_c}$, it is isomorphic to a general WCI $Y_{d_1,\ldots,d_{c'}} \subset \mathbb{P}(a_{c_1}, \ldots, a_n)$.

Thus the base locus is again a WCI. However this is not necessarily (quasi)smooth in general. We shall see this in Example 5.6.

Example 5.6. Let $X := X_{231,231,26} \subset \mathbb{P} := \mathbb{P}(3, 3, 7, 7, 11, 11, 1^{(447)})$ be a general WCI. We can check that this is a smooth Fano WCI as follows: for $I = \{0, 1\}, \{2, 3\}$ or $\{4, 5\}$, (that is, two variables of weights 3, 7 or 11), we have (Q1) for $d_1 = 231$, $d_2 = 231$. Also, for $I = \{0, 1, 2, 3\}$ or $\{0, 1, 4, 5\}$, we have (Q1) for $d_1 = 231$, $d_2 = 231$, $d_3 = 26$ since $26 = 7 \cdot 2 + 3 \cdot 4 = 11 + 3 \cdot 5$. For $I = \{2, 3, 4, 5\}$, we have (Q2) for $d_1 = 231$, $d_2 = 231$, $d_3 = 26 = 7 \cdot 2 + 11 + 1$. By Proposition 3.1, we see that X is quasismooth, and smooth since $X \cap \operatorname{Sing} \mathbb{P} = \varnothing$.

The base locus $\operatorname{Bs} |\mathcal{O}_X(1)|$ is a WCI $Y := Y_{231,231,26} \subset \mathbb{P}' := \mathbb{P}(3, 3, 7, 7, 11, 11)$. This is not quasismooth. Indeed, for $I = \{2, 3, 4, 5\}$, neither (Q1) nor (Q2) holds because of the lack of suitable degree 26 polynomials. In fact, Y is a nonnormal surface singular along a curve $(x_0 = x_1 = f_1 = f_2 = 0) \subset \mathbb{P}'$, where f_1, f_2 are part of defining polynomials of degrees 231 and x_0, x_1 are the variables of weights 3.

Hence we can not expect smoothness of the base locus of the fundamental linear system even if it contains a smooth member.

Remark 5.7. Let $W = W_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a smooth WCI which is not a linear cone, where $a_i > 1$ for any $i = 0, \dots, n$. By Corollary 5.3 we know that W is not Fano. Then we can consider a WCI

$$X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n, 1^{(\ell)})$$

where $\ell = \delta(W) + 1$. In this way X is a smooth Fano with $-K_X = \mathcal{O}_X(1)$ and $\operatorname{Bs} |\mathcal{O}_X(1)|$ is exactly W.

In Corollary 5.3 we showed that for a smooth Fano WCI, the general member of the fundamental divisor is quasismooth. This is not true in general for a quasismooth Fano WCI as the following example shows.

Example 5.8. Let $X = X_{35} \subset \mathbb{P}(5, 7, 2^{(k)}, 3^{(k)})$ where $k \geq 5$. Then X is a quasismooth Fano WCI with fundamental divisor $\mathcal{O}_X(6)$, but $X_{35,6} \subset \mathbb{P}(5, 7, 2^{(k)}, 3^{(k)})$ is not quasismooth. However, we see that a general member of $|\mathcal{O}_X(6)|$ has only terminal singularities. Indeed it has an isolated singularity at $[* : * : 0 : \cdots : 0]$ which is locally isomorphic to $0 \in (x_1^3 + \cdots + x_k^3 + x_{k+1}^2 + \cdots + x_{2k}^2 = 0) \subset \mathbb{C}^{2k}$.

It is also natural to look at the general element of $|-K_X|$ in the case of a Fano variety X. For instance, Shokurov [1979] and Reid [1983] proved that a Fano 3-fold with only canonical Gorenstein singularities admits an anticanonical member with only Du Val singularities. Here we give an example of a smooth Fano WCI whose anticanonical members are singular (and not quasismooth). See also [Höring and Voisin 2011, 2.12] for an example of a Fano 4-fold with singular fundamental divisor.

Example 5.9. (cf., [Sano 2014, Example 2.9]) For $m \in \mathbb{Z}_{>0}$, let X be a weighted hypersurface $X = X_{(2m+1)(2m+2)} \subset \mathbb{P}(1^{(1+2m(2m+1))}, 2m+1, 2m+2)$ of degree $(2m+1)(2m+2)$. Then we see that $-K_X = \mathcal{O}_X(2)$ and the linear system $|-iK_X|$ does not contain a smooth member for $i = 1, \dots, m$. Thus we can not expect a smooth element of the plurianticanonical system on a Fano manifold. However, in the above example, we can find a member with only terminal singularities. Moreover, the base locus of $|H|$ consists of a point.

Remark 5.10. It is well known that following the arguments in [Ambro 1999, Section 5] or [Kawamata 2000, Section 5] and assuming Conjecture 1.1, it is possible to show that the general element of $|-mK_X|$ has always only klt singularities for $m > 0$ such that $-mK_X$ is Cartier (we thank Chen Jiang for pointing this fact out to us).

Finally, we also get the following corollary, which generalizes [Przyjalkowski and Shramov 2017, Corollary 4.2] to any codimension.

Corollary 5.11. *Let $X := X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a well-formed smooth Fano or Calabi–Yau WCI which is not a linear cone. Let $c_1 := |\{i \in \bar{n} : a_i = 1\}|$. Then $c_1 > I(X) := -\delta(d; a) = \sum_{i=0}^n a_i - \sum_{j=1}^c d_j$.*

Proof. Consider the regular pair $(d; a)$ associated with X. We may assume that $a_0 \leq \cdots \leq a_n$, so that $a_0 = \cdots = a_{c_1-1} = 1$. Let $(d'; a')$ be the pair $(d; a_{c_1}, \dots, a_n)$, where we took away every 1 from $(d; a)$. This pair is regular with no $a_i = 1$ and so by Proposition 5.2 we get

$$\delta(d'; a') \geq c > 0,$$

which implies

$$\delta(d; a) = \delta(d'; a') - c_1 > -c_1,$$

i.e., $I(X) < c_1$, as we wanted. □

General case. The following is a key proposition to deduce the nonvanishing in the quasismooth Fano case.

Proposition 5.12. *Let $h \in \mathbb{N}_+$ and $(d; a) \in \mathbb{N}_+^c \times \mathbb{N}_+^{n+1}$ be an h-regular pair with $c \geq 1$. If $a_i \nmid h$ for any $i = 0, \dots, n$ and $a_i \neq d_j$ for any i, j, then*

$$\delta(d; a) > 0.$$

Proof. Let us write $h = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, where the p_i are distinct prime numbers. The proof is by induction on $\alpha = \sum_{i=1}^k \alpha_i \geq 0$. If $\alpha = 0$, then the pair $(d; a)$ is regular and the statement follows from Proposition 5.2, so we assume $\alpha \geq 1$.

Let p be a prime number dividing h and consider $(d^p; a^p)$. As usual,

$$\delta(d; a) = \delta(d^p; a^p) + \frac{p-1}{p}\delta(d(p); a(p)).$$

By Lemma 4.6, $(d^p; a^p)$ and $(d(p)/p; a(p)/p)$ are h/p-regular. Note that there does not exist i such that $a_i/p = 1$ by the hypothesis $a_i \nmid h$. Thus, after cancellation on $(d^p; a^p)$ (see Definition 4.3), we see that $(\tilde{d^p}; \tilde{a^p})$ and $(d(p)/p; a(p)/p)$ are h/p-regular and satisfy the hypothesis of the proposition.

If $p \mid a_i$ or $p \mid d_j$ for some i or j, then we obtain $\delta(d(p)/p; a(p)/p) > 0$ by the induction hypothesis and conclude $\delta(d; a) > 0$ by induction since we have either $\delta(d^p; a^p) = \delta(\tilde{d^p}; \tilde{a^p}) > 0$ or $(\tilde{d^p}; \tilde{a^p})$ is empty.

If $p \nmid a_i$ and $p \nmid d_j$ for any i, j, then $\delta(d^p; a^p) > 0$ by the induction hypothesis since $(d^p; a^p) = (d; a)$ is h/p-regular. Moreover $(d(p); a(p))$ is empty. Hence we can again conclude that $\delta(d; a) > 0$. □

Corollary 5.13. *Let $X = X_{d_1,\dots,d_c} \subset \mathbb{P}(a_0, \dots, a_n)$ be a well-formed quasismooth WCI which is Fano or Calabi–Yau and which is not a linear cone. Then $|H| \neq \varnothing$ for any ample Cartier divisor H on X.*

Proof. Write $H = \mathcal{O}_X(h)$. If there exists $i \in \bar{n}$ such that $a_i \mid h$, then we are done. Otherwise, we are in the situation of Proposition 5.12, and so the variety can not be Fano or Calabi–Yau since $(d; a)$ is h-regular by Proposition 3.6. □

6. Weighted hypersurfaces

The following lemma gives a proof of a generalized version of Conjecture 4.8 in the case $c = 1$.

Lemma 6.1. *Let $a_0, \dots, a_n$ be positive integers, $n \geq 1$ and set*

$$h := \mathrm{lcm}_{i \neq j}(\gcd(a_i, a_j)).$$

Assume that $a_i \nmid h$ for any i and set $f := \mathrm{lcm}(a_0, \dots, a_n)$. Then

$$f - \sum_{i=0}^{n} a_i \geq \mathrm{lcm}(a_s, a_t) - a_s - a_t$$

for any s and t.

Proof. We first note that, for any proper subset I of $\bar{n} := \{0, \dots, n\}$, we have $f \neq \mathrm{lcm}_{i \in I}(a_i)$. In fact, suppose that the equality holds and let $k \in \bar{n} \setminus I$. For any prime power p^e such that $e \geq 1$ and $p^e \mid a_k$, we have $p^e \mid f$. In particular, we have $p^e \mid a_\ell$ for some $\ell \in I$. This implies that $p^e \mid \gcd(a_k, a_\ell)$ and so $a_k \mid h$, which is a contradiction. In particular, $f \geq 2\,\mathrm{lcm}_{i \in I}(a_i)$.

The proof of the lemma is by induction on n and the case $n = 1$ is trivial, so assume $n \geq 2$. Let $s, t \in \bar{n}$ be such that $s \neq t$. Then

$$f - \sum_{i=0}^{n} a_i = \frac{f}{2} - a_s - a_t + \frac{f}{2} - \sum_{i \neq s,t} a_i \geq \mathrm{lcm}(a_s, a_t) - a_s - a_t + \mathrm{lcm}_{i \neq s,t}(a_i) - \sum_{i \neq s,t} a_i.$$

If $n=2$, then we have

$$\operatorname{lcm}_{i\neq s,t}(a_i)-\sum_{i\neq s,t}a_i=0$$

and we are done.

If $n\geq 3$, then we have

$$\operatorname{lcm}_{i\neq s,t}(a_i)-\sum_{i\neq s,t}a_i\geq \operatorname{lcm}(a_{s'},a_{t'})-a_{s'}-a_{t'}$$

for $s',t'\in\bar{n}\setminus\{s,t\}$ such that $s'\neq t'$ by induction on n and $\operatorname{lcm}(a_{s'},a_{t'})-a_{s'}-a_{t'}\geq 0$ because we are assuming that $a_i\nmid h$ for any i. □

Proposition 6.2. *Let $X=X_d\subset\mathbb{P}=\mathbb{P}(a_0,\dots,a_n)$ be a well-formed, quasismooth hypersurface of degree d which is not a linear cone. Let H be an ample Cartier divisor on X such that $H-K_X$ is ample.*

Then $|H|$ is not empty.

Proof. Write $\mathcal{O}_X(H)=\mathcal{O}_X(h)$ for a positive integer h.

By Proposition 3.4, we can assume that $a_i\mid d$ for any i. Then X is a Cartier divisor which intersects any stratum $P_{\{i,j\}}$ in some interior point. The condition of H to be Cartier is then equivalent to

$$\operatorname{lcm}_{i\neq j}(\gcd(a_i,a_j))\mid h.$$

If there exists a_i such that $a_i\mid h$, then we are done. So assume that $a_i\nmid h$ for any i and let $f:=\operatorname{lcm}(a_0,\dots,a_n)$. By Lemma 6.1, we get

$$f-\sum_{i=0}^{n}a_i\geq \operatorname{lcm}(a_s,a_t)-a_s-a_t$$

for any s and t. Since $h>f-\sum_{i=0}^{n}a_i$ (because $H-K_X$ is ample and $f\mid d$) and $g:=\gcd(a_s,a_t)\mid h$ for any $s\neq t$, we can use the Frobenius number $G(a_s/g,a_t/g)=(1/g)(\operatorname{lcm}(a_s,a_t)-a_s-a_t)$ as on page 2382 to conclude that there are nonnegative integers λ_s,λ_t such that

$$\lambda_s a_s+\lambda_t a_t=h,$$

which implies that $|H|$ is not empty by Lemma 2.4. □

In the following, we prove the basepoint freeness on a Gorenstein weighted hypersurface.

Proposition 6.3. *Let $X=X_d\subset\mathbb{P}=\mathbb{P}(a_0,\dots,a_n)$ be a well-formed, quasismooth hypersurface of degree d which is not a linear cone such that K_X is Cartier. Let H be the fundamental divisor of X and h be the positive integer such that $H=\mathcal{O}_X(h)$.*

Then $L=K_X+mH$ is globally generated for any $m\geq n$.

Proof. Suppose by contradiction that there is a point $p = [p_0 : \cdots : p_n] \in \mathrm{Bs}\,|L|$ and take ℓ such that $L = \mathcal{O}_X(\ell)$.

Note that if $p_s \neq 0$ for some s, then $a_s \nmid h$, otherwise $x_s^e \in |L|$ for some positive integer e and so $p \notin \mathrm{Bs}\,|L|$. Also note that, for all $i \in \bar{n}$ such that $a_i \nmid h$, we have $a_i \mid d$ by Proposition 3.4.

Assume first that there exists a unique $s \in \bar{n}$ such that $p_s \neq 0$. Since $p \in X$ and $a_s \nmid h$, we get that $a_s \mid d$. Let f_d be the defining polynomial of X_d. If f_d contains a monomial x_s^{d/a_s}, then we obtain $p \notin X_d$ and this is a contradiction. If f_d does not contain such a monomial, then it should contain a monomial of the form $x_s^k x_i$ for some $k > 0$ and $i \neq s$ by the quasismoothness of X_d. Then we see that $a_s \mid a_i$ by $a_s \mid d$, and X_d has a quotient singularity of index a_s. Thus we obtain $a_s \mid h$ and this is a contradiction.

Hence we can assume that there exist s and t such that $s \neq t$, $p_s \neq 0$ and $p_t \neq 0$, thus $a_s, a_t \nmid h$. We have

$$\ell = d - \sum_{i=0}^{n} a_i + mh = d - \sum_{a_i \nmid h} a_i - \sum_{a_i \mid h} a_i + mh.$$

Assume that $-\sum_{a_i \mid h} a_i + mh \geq 1$. Since $a_i \mid d$ for all i such that $a_i \nmid h$, we can apply Lemma 6.1 to conclude that

$$\ell > \mathrm{lcm}(a_s, a_t) - a_s - a_t,$$

which implies that $x_s^{e_s} x_t^{e_t} \in |L|$ for some nonnegative integers e_s and e_t. So we again have $p \notin \mathrm{Bs}\,|L|$.

Assume now that $-\sum_{a_i \mid h} a_i + mh \leq 0$. Then we can check that $|\{i : a_i = h\}| \geq n-1$, because $m \geq n$. Moreover, since $\mathbb{P}$ is well-formed, the greatest common factor of any n weights is 1. By these, when $|\{i : a_i = h\}| = n$, we have $h = 1$ and $\mathbb{P} = \mathbb{P}(a_0, 1, \ldots, 1)$ for some $a_0 > 1$. When $|\{i : a_i = h\}| = n-1$, we have $h = 2$ and $\mathbb{P} = \mathbb{P}(1, 1, 2, \ldots, 2)$. In both cases, we can check that L is basepoint free, and we have derived a contradiction. □

Acknowledgment

We thank D. Turchetti for many helpful conversations, C. Shramov for letting us know of the recent preprint [Przyjalkowski and Shramov 2017], and A. Höring for asking us a question on the base locus of the fundamental linear system. Part of this project was realized while the Sano and Tasin were visiting the Max Planck Institute for Mathematics and the Institute for Mathematical Science, NUS in Singapore. We are happy to thank them for their support during our stays. Sano was also partially supported by JST Tenure Track Program, JSPS KAKENHI Grant Numbers JP15J03158, JP16K17573, JP24224001.

References

[Ambro 1999] F. Ambro, "Ladders on Fano varieties", *J. Math. Sci.* (*New York*) **94**:1 (1999), 1126–1135. MR Zbl

[Ballico et al. 2013] E. Ballico, R. Pignatelli, and L. Tasin, "Weighted Hypersurfaces with Either Assigned Volume or Many Vanishing Plurigenera", *Communications in Algebra* **41**:10 (2013), 3745–3752.

[Beltrametti and Sommese 1995] M. C. Beltrametti and A. J. Sommese, *The adjunction theory of complex projective varieties*, de Gruyter Expositions in Mathematics **16**, de Gruyter, Berlin, 1995. MR Zbl

[Brauer 1942] A. Brauer, "On a problem of partitions", *Amer. J. Math.* **64** (1942), 299–312. MR Zbl

[Brauer and Shockley 1962] A. Brauer and J. E. Shockley, "On a problem of Frobenius", *J. Reine Angew. Math.* **211** (1962), 215–220. MR Zbl

[Broustet and Höring 2010] A. Broustet and A. Höring, "Effective non-vanishing conjectures for projective threefolds", *Adv. Geom.* **10**:4 (2010), 737–746. MR Zbl

[Call and Lyubeznik 1994] F. Call and G. Lyubeznik, "A simple proof of Grothendieck's theorem on the parafactoriality of local rings", pp. 15–18 in *Commutative algebra: syzygies, multiplicities, and birational algebra* (South Hadley, MA, 1992), edited by W. J. Heinzer et al., Contemp. Math. **159**, American Mathematical Society, Providence, RI, 1994. MR Zbl

[Cao and Jiang 2016] Y. Cao and C. Jiang, "Remarks on Kawamata's effective non-vanishing conjecture for manifolds with trivial first Chern classes", preprint, 2016. arXiv

[Chen 2015] J.-J. Chen, "Finiteness of Calabi–Yau quasismooth weighted complete intersections", *Int. Math. Res. Not.* **2015**:12 (2015), 3793–3809. MR Zbl

[Chen et al. 2011] J.-J. Chen, J. A. Chen, and M. Chen, "On quasismooth weighted complete intersections", *J. Algebraic Geom.* **20**:2 (2011), 239–262. MR Zbl

[Corti et al. 2000] A. Corti, A. Pukhlikov, and M. Reid, "Fano 3-fold hypersurfaces", pp. 175–258 in *Explicit birational geometry of* 3*-folds*, edited by A. Corti and M. Reid, London Math. Soc. Lecture Note Ser. **281**, Cambridge Univ. Press, 2000. MR Zbl

[Dimca 1986] A. Dimca, "Singularities and coverings of weighted complete intersections", *J. Reine Angew. Math.* **366** (1986), 184–193. MR Zbl

[Dolgachev 1982] I. Dolgachev, "Weighted projective varieties", pp. 34–71 in *Group actions and vector fields* (Vancouver, 1981), edited by J. B. Carrell, Lecture Notes in Math. **956**, Springer, 1982. MR Zbl

[Ein and Lazarsfeld 1993] L. Ein and R. Lazarsfeld, "Global generation of pluricanonical and adjoint linear series on smooth projective threefolds", *J. Amer. Math. Soc.* **6**:4 (1993), 875–903. MR Zbl

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies **131**, Princeton Univ. Press, 1993. MR Zbl

[Höring 2012] A. Höring, "On a conjecture of Beltrametti and Sommese", *J. Algebraic Geom.* **21**:4 (2012), 721–751. MR Zbl

[Höring and Voisin 2011] A. Höring and C. Voisin, "Anticanonical divisors and curve classes on Fano manifolds", *Pure Appl. Math. Q.* **7**:4 (2011), 1371–1393. MR Zbl

[Iano-Fletcher 2000] A. R. Iano-Fletcher, "Working with weighted complete intersections", pp. 101–173 in *Explicit birational geometry of* 3*-folds*, edited by A. Corti and M. Reid, London Math. Soc. Lecture Note Ser. **281**, Cambridge Univ. Press, 2000. MR Zbl

[Johnson 1960] S. M. Johnson, "A linear diophantine problem", *Canad. J. Math.* **12** (1960), 390–398. MR Zbl

[Johnson and Kollár 2001] J. M. Johnson and J. Kollár, "Fano hypersurfaces in weighted projective 4-spaces", *Experiment. Math.* **10**:1 (2001), 151–158. MR Zbl

[Kawamata 1997] Y. Kawamata, "On Fujita's freeness conjecture for 3-folds and 4-folds", *Math. Ann.* **308**:3 (1997), 491–505. MR Zbl

[Kawamata 2000] Y. Kawamata, "On effective non-vanishing and base-point-freeness", *Asian J. Math.* **4**:1 (2000), 173–181. MR Zbl

[Lanteri et al. 1993] Lanteri et al., "Open problems", pp. 321–325 in *Geometry of complex projective varieties* (Cetraro, Italy, 1990), edited by A. Lanteri et al., Seminars and Conferences **9**, Mediterranean Press, Rende, Italy, 1993. MR

[Mori 1975] S. Mori, "On a generalization of complete intersections", *J. Math. Kyoto Univ.* **15**:3 (1975), 619–646. MR Zbl

[Przyjalkowski and Shramov 2016] V. Przyjalkowski and C. Shramov, "Bounds for smooth Fano weighted complete intersections", preprint, 2016. arXiv

[Przyjalkowski and Shramov 2017] V. Przyjalkowski and C. Shramov, "Nef partitions for codimension 2 weighted complete intersections", preprint, 2017. arXiv

[Reid 1980] M. Reid, "Canonical 3-folds", pp. 273–310 in *Journées de Géometrie Algébrique* (Angers, 1979), edited by A. Beauville, Sijthoff & Noordhoff, Germantown, MD, 1980. MR Zbl

[Reid 1983] M. Reid, "Projective morphisms according to Kawamata", preprint, 1983, Available at http://homepages.warwick.ac.uk/~masda/3folds/Ka.pdf.

[Reid 1987] M. Reid, "Young person's guide to canonical singularities", pp. 345–414 in *Algebraic geometry, I* (Brunswick, ME, 1985), edited by S. J. Bloch, Proc. Sympos. Pure Math. **46**, American Mathematical Society, Providence, RI, 1987. MR Zbl

[Reider 1988] I. Reider, "Vector bundles of rank 2 and linear systems on algebraic surfaces", *Ann. of Math.* (2) **127**:2 (1988), 309–316. MR Zbl

[Sano 2014] T. Sano, "Unobstructedness of deformations of weak Fano manifolds", *Int. Math. Res. Not.* **2014**:18 (2014), 5124–5133. MR Zbl

[Shokurov 1979] V. V. Shokurov, "Smoothness of a general anticanonical divisor on a Fano variety", *Izv. Akad. Nauk SSSR Ser. Mat.* **43**:2 (1979), 430–441. In Russian; translation in *Math. USSR-Izv.* **14**:2 (1980), 395–405. MR Zbl

[Xie 2009] Q. Xie, "A note on the effective non-vanishing conjecture", *Proc. Amer. Math. Soc.* **137**:1 (2009), 61–63. MR Zbl

[Ye and Zhu 2015] F. Ye and Z. Zhu, "On Fujita's freeness conjecture in dimension 5", preprint, 2015. arXiv

Communicated by Shigefumi Mori
Received 2017-03-29 Revised 2017-07-28 Accepted 2017-09-01

marco.pizzato1@gmail.com *I-39100 Bolzano, Italy*

tarosano@math.kobe-u.ac.jp *Department of Mathematics, Graduate School of Science, Kobe University, Kobe, Japan*

ltasin@mat.uniroma3.it *Dipartimento di Matematica e Fisica, Roma Tre University, Rome, Italy*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2397

Generalized Kuga–Satake theory and good reduction properties of Galois representations

Stefan Patrikis

In previous work, we described conditions under which a single geometric representation $\Gamma_F \to H(\overline{\mathbb{Q}}_\ell)$ of the Galois group of a number field F lifts through a central torus quotient $\widetilde{H} \to H$ to a geometric representation. In this paper, we prove a much sharper result for systems of ℓ-adic representations, such as the ℓ-adic realizations of a motive over F, having common "good reduction" properties. Namely, such systems admit geometric lifts with good reduction outside a common finite set of primes. The method yields new proofs of theorems of Tate (the original result on lifting projective representations over number fields) and Wintenberger (an analogue of our main result in the case of a central isogeny $\widetilde{H} \to H$).

1. Introduction

Let F be a number field, and let $\Gamma_F = \mathrm{Gal}(\overline{F}/F)$ be its absolute Galois group with respect to a fixed algebraic closure $\overline{F}$. A fundamental theorem of Tate (see [Serre 1977, §6]) asserts that $H^2(\Gamma_F, \mathbb{Q}/\mathbb{Z})$ vanishes; as a result, all (continuous, ℓ-adic) projective representations of Γ_F lift to genuine representations, and more generally, whenever $\widetilde{H} \to H$ is a surjection of linear algebraic groups over $\overline{\mathbb{Q}}_\ell$ with kernel equal to a central torus in $\widetilde{H}$, all representations $\rho_\ell : \Gamma_F \to H(\overline{\mathbb{Q}}_\ell)$ lift to $\widetilde{H}(\overline{\mathbb{Q}}_\ell)$.

The ℓ-adic representations of greatest interest in number theory are those with conjectural connections to the theories of motives and automorphic forms; if the monodromy group of ρ_ℓ is semisimple, then it is expected — by conjectures of Fontaine–Mazur, Tate, Grothendieck–Serre, and Langlands — that the ρ_ℓ arising from pure motives or automorphic forms are precisely those that are geometric in the sense of Fontaine–Mazur, i.e., unramified outside a finite set of places of F, and de Rham at all places dividing ℓ. The paper [Patrikis 2016c] established a variant of Tate's lifting theorem for such geometric Galois representations. There are

I am grateful to the anonymous referee for helpful comments that have improved the exposition.
MSC2010: primary 11F80; secondary 11R37.
Keywords: Galois representations, Kuga–Satake construction.

obstructions when F has real embeddings, but at least for totally imaginary F, any geometric $\rho_\ell : \Gamma_F \to H(\overline{\mathbb{Q}}_\ell)$ satisfying a natural "Hodge symmetry" requirement admits a geometric lift $\tilde{\rho}_\ell : \Gamma_F \to \widetilde{H}(\overline{\mathbb{Q}}_\ell)$ [Patrikis 2016c, Theorem 3.2.10]. This geometric lifting theorem leads to a precise expectation for the corresponding lifting problem for motivic Galois representations. Namely, if $\mathcal{G}_{F,E}$ denotes the motivic Galois group for pure motives over F with coefficients in a number field E — we will make this setup precise in Section 2, but for now the reader may take homological motives under the standard conjectures — and if $\widetilde{H} \to H$ is now a surjection of groups over E with central torus kernel, then we conjecture [Patrikis 2016c, Conjecture 4.3.1] that any motivic Galois representation $\rho : \mathcal{G}_{F,E} \to H$ lifts to $\widetilde{H}$, at least after some finite extension of coefficients:

$$\begin{tikzcd} & \widetilde{H}_{\overline{E}} \arrow[d] \\ \mathcal{G}_{F,\overline{E}} \arrow[ur, dashed, "\tilde{\rho}"] \arrow[r, "\rho\otimes_E \overline{E}"'] & H_{\overline{E}} \end{tikzcd}$$

There is essentially one classical example (with several variants) of this conjecture, a well-known construction of Kuga and Satake [1967], which associates to a complex, for our purposes projective, K3 surface X a complex abelian variety $\mathrm{KS}(X)$, related by an inclusion of Hodge-structures $H^2(X, \mathbb{Q}) \subset H^1(\mathrm{KS}(X), \mathbb{Q})^{\otimes 2}$. In the motivic Galois language, finding $\mathrm{KS}(X)$ amounts (when $F = \mathbb{C}$) to finding a lift $\tilde{\rho}$ of the representation $\rho_X : \mathcal{G}_{\mathbb{C},\mathbb{Q}} \to H = \mathrm{SO}(H^2(X)(1))$, through the surjection $\widetilde{H} = \mathrm{GSpin}(H^2(X)(1)) \to H$. Progress on the general conjecture, when the motives in question do not lie in the Tannakian subcategory of motives generated by abelian varieties, seems to require entirely new ideas.[1]

The aim of this paper is to establish a Galois-theoretic result which is necessary for this conjecture to hold, but considerably more delicate than the basic geometric lifting theorem of [Patrikis 2016c, Theorem 3.2.10]. Namely, any motive M over F has good reduction outside a finite set of primes: for any choice of variety X in whose cohomology M appears, X spreads out as a smooth projective scheme over $\mathcal{O}_F[1/N]$ for some integer N. In particular, by the base-change theorems of étale cohomology [Deligne 1977] and the crystalline p-adic comparison isomorphism [Faltings 1989], for any motivic Galois representation $\rho : \mathcal{G}_{F,E} \to H$, the λ-adic realizations $\rho_\lambda : \Gamma_F \to H(E_\lambda)$ have *good reduction* outside a finite set of primes S, in the sense (also see Definition 1.1) that each ρ_λ factors through $\Gamma_{F,S\cup S_\lambda}$ and is crystalline at all places of $S_\lambda \setminus (S_\lambda \cap S)$; here S_λ denotes the primes of F with the same residue characteristic as λ, and $\Gamma_{F,S\cup S_\lambda}$ is the Galois group of the maximal extension of F inside $\overline{F}$ that is unramified outside of $S \cup S_\lambda$. Certainly a necessary

[1] For some, admittedly limited, examples, see [Patrikis 2016a; 2016b].

condition for the generalized Kuga–Satake conjecture to hold is that the realizations $\{\rho_\lambda\}_\lambda$ of ρ should lift to geometric representations $\tilde{\rho}_\lambda : \Gamma_{F,P\cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda)$ that likewise have good reduction outside a common finite set of places P. This is what we will show, as a consequence of a more general result. To state it, we first make a couple of definitions.

Definition 1.1. A collection $\{\rho_\lambda : \Gamma_F \to H(\overline{E}_\lambda)\}_\lambda$, as λ varies over finite places of E, of geometric Galois representations is *ramification-compatible* if there exist

(1) a finite set S of places of F such that each ρ_λ is unramified outside of $S \cup S_\lambda$, i.e., factors through
$$\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(\overline{E}_\lambda),$$
and for v in S_λ but not in S, $\rho_\lambda|_{\Gamma_{F_v}}$ is crystalline; and

(2) a central cocharacter $\omega : \boldsymbol{G}_{m,E} \to H$ and a collection of conjugacy classes
$$\{[\mu_\tau : \boldsymbol{G}_{m,\overline{E}} \to H_{\overline{E}}]\}_{\tau : F \hookrightarrow \overline{E}}$$
satisfying $[\mu_\tau] = \omega\cdot[\mu_{c\tau}^{-1}]$ for any choice of complex conjugation $c \in \mathrm{Gal}(\overline{E}/\mathbb{Q})$, such that for all E-embeddings $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$, inducing via τ some $\tau_{\iota_\lambda} : F_v \hookrightarrow \overline{E}_\lambda$, the conjugacy class $[\mu_\tau \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda]$ is equal to the conjugacy class of τ_{ι_λ}-labeled Hodge–Tate cocharacters associated to $\rho_\lambda|_{\Gamma_{F_v}}$.

If a single representation ρ_λ satisfies the condition in item (1), we say ρ_λ *has good reduction outside* S. If it satisfies the condition in item (2) (for some collection of cocharacters ω, $\{\mu_\tau\}$), then we say it satisfies *Hodge symmetry*.

Remark 1.2.

- Note that the ρ_λ need not be "compatible" in the usual sense (frobenii acting compatibly): if the "coefficients" of the ρ_λ are bounded in a rather strong sense — there exists a common number field over which their frobenius characteristic polynomials are defined — one expects that our collection of ρ_λ should partition (dividing up the λ's) into finitely many compatible systems.[2]
- The Hodge symmetry requirement of part (2) of Definition 1.1 is not the most general constraint that pertains to a compatible system of ℓ-adic representations. It will always hold for the λ-adic realizations of motives, as we will see in Lemma 2.3, when we take the ambient group to be the motivic Galois group G_ρ of the underlying motivic Galois representation ρ, and not some larger group. But there may be compatible systems (of motivic origin) where the criterion in part (2) of Definition 1.1 fails; for example, consider $\rho_\ell : \Gamma_{\mathbb{Q}} \to \mathrm{PGL}_3(\overline{\mathbb{Q}}_\ell)$ given

[2] To see the relevance of bounding the coefficients, the reader may contrast the case of elliptic curves (over $\mathbb{Q}$, say) unramified outside S with that of all weight-two modular forms unramified outside S: of the former there are finitely many isogeny classes, since the conductor is bounded, whereas the latter can have level divisible by arbitrarily high powers of the primes in S.

by the projectivization of $\kappa_\ell \oplus 1 \oplus 1$, with κ_ℓ denoting the ℓ-adic cyclotomic character. The relation $[\mu_\tau] = \omega \cdot [\mu_{c\tau}^{-1}]$ implies in this case $[\mu_\tau] = [\mu_\tau^{-1}]$, which is false. A way around this is given in [Patrikis 2016c, §3.2], where the Hodge symmetry hypothesis is formulated in a way that conjecturally holds for any geometric representation $\rho_\ell : \Gamma_F \to H(\overline{\mathbb{Q}}_\ell)$, regardless of its (reductive) algebraic monodromy group. The proof of [Patrikis 2016c, Theorem 3.2.10] thus requires a slightly trickier group-theoretic argument than the one we require here. We have opted in this paper to keep the simpler condition (2) above, so as to focus on what is new in the arguments, and because of its obvious centrality from a motivic point of view (in particular, its sufficiency for Corollary 1.6).

Here is the main theorem:

Theorem 1.3. *Let E be a number field, and let $\widetilde{H} \to H$ be a surjection of linear algebraic groups over E with kernel equal to a central torus in $\widetilde{H}$. Let F be a totally imaginary number field, and let S be a finite set of places of F containing the archimedean places. Fix a set of cocharacters $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ satisfying the "Hodge symmetry" condition of part (2) of Definition 1.1. Then there exists a finite set of places $P \supset S$ such that for any place λ of E, any embedding $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$, and any geometric representation $\rho_\lambda : \Gamma_{F,S \cup S_\lambda} \to H(\overline{E}_\lambda)$ such that*

- *ρ_λ has good reduction outside S, and*
- *the conjugacy classes of labeled Hodge–Tate cocharacters of ρ_λ are induced via ι_λ from $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ (again, see Definition 1.1 for details),*

the representation ρ_λ admits a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F,P \cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda)$ having good reduction outside P.

In particular, if $\{\rho_\lambda : \Gamma_{F,S \cup S_\lambda} \to H(\overline{E}_\lambda)\}_\lambda$ is a ramification-compatible system, then there exist a finite set of places $P \supset S$ and lifts $\tilde{\rho}_\lambda : \Gamma_{F,P \cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda)$ such that $\{\tilde{\rho}_\lambda\}_\lambda$ is a ramification-compatible system.

Remark 1.4. All results of this paper, once we take into account the caveat of [Patrikis 2016c, §2.8] (see too [Patrikis 2015, Proposition 5.5]), admit straightforward variants when F has real places. Thus, for real F, the analogue of Theorem 1.3 either holds exactly as written, or after replacing F by any totally imaginary (e.g., composite with a quadratic imaginary) extension. We do not want to discuss this at any length here, but we simply remind the reader that the prototypical example in which F is totally real, and Theorem 1.3 fails as stated, is that of the projective motivic Galois representation associated to a *mixed-parity* Hilbert modular form.

The proof of this theorem is completed in Theorem 1.3. The typical application is to the collection of Galois representations $\{\rho_\lambda\}_\lambda$ associated to a motivic Galois representation; we make this precise in Corollary 1.6.

Next we describe applications of Theorem 1.3 to the more general problem of lifting through surjections $H' \to H$ with central kernel of multiplicative type. Here in general we cannot expect as strong a result as Theorem 1.3. First, the Hodge–Tate cocharacters of ρ_λ may not lift to H', in which case there can be no geometric lifts to H'. Second, even if the Hodge–Tate cocharacters lift, the Galois representations may only lift after a finite base change of F. For example, if ρ_λ is the projectivization of the Galois representation associated to a weight 3 modular form, $\det(\rho_\lambda) : \Gamma_{\mathbb{Q}} \to \overline{E}_\lambda^\times$ does not admit a square root. A beautiful result of Wintenberger [1995, Théorème 2.1.4, Théorème 2.1.7] shows that when $H' \to H$ is an isogeny, a result similar to Theorem 1.3 holds, as long as in the conclusion F is replaced by a suitable finite extension. Here we treat the general case of multiplicative-type quotients:

Theorem 1.5 (see Corollary 3.18). *Let $H' \to H$ be a surjection of linear algebraic groups over E whose kernel is central and of multiplicative type. Let F be a number field, and let S be a finite set of places of F containing the archimedean places. Fix a set of cocharacters $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ as in part (2) of Definition 1.1, and moreover, assume that each μ_τ lifts to a cocharacter of H'.*

Then there exist a finite set of places $P \supset S$, and a finite extension F'/F, such that any geometric representation $\rho_\lambda : \Gamma_{F, S \cup S_\lambda} \to H(\overline{E}_\lambda)$ having good reduction outside S, and whose Hodge–Tate cocharacters arise from the set $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ via some embedding $\overline{E} \hookrightarrow \overline{E}_\lambda$, admits a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F', P \cup S_\lambda} \to H'(\overline{E}_\lambda)$ having good reduction outside P.

In particular, if $\{\rho_\lambda : \Gamma_{F, S \cup S_\lambda} \to H(\overline{E}_\lambda)\}_\lambda$ is a ramification-compatible system with Hodge cocharacters $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$, then there exist a finite set of places $P \supset S$, a finite extension F'/F, and lifts $\tilde{\rho}_\lambda : \Gamma_{F', P \cup S_\lambda} \to H'(\overline{E}_\lambda)$ such that $\{\tilde{\rho}_\lambda\}_\lambda$ is a ramification-compatible system.

We deduce Wintenberger's original result in Corollary 3.16. Our proof differs in an essential way, as it passes through Theorem 1.3, which cannot be deduced from the methods of [Wintenberger 1995]. Our problem resembles Wintenberger's in that both lead to a basic difficulty of annihilating cohomological obstruction classes in infinitely many Galois cohomology groups, one for each λ, but needing to do so in an "independent-of-λ" fashion. The arguments themselves, however, are in fact orthogonal to one another: Wintenberger always kills cohomology by making a finite base change on F, whereas that is precisely what we are forbidden from doing if we want the more precise results of Theorem 1.3. Moreover, our methods also yield a novel proof of Tate's original vanishing theorem (see Corollary 3.9). In fact, Corollary 3.9 establishes a more precise form of Tate's theorem: the latter of course shows that the image under the canonical map $H^2(\Gamma_{F,S}, \mathbb{Z}/N) \to H^2(\Gamma_F, \mathbb{Q}/\mathbb{Z})$ is zero, and our refinement quantifies how much additional ramification must be added, and how much the coefficients must be enlarged, in order to annihilate $H^2(\Gamma_{F,S}, \mathbb{Z}/N)$. Our arguments thus achieve, from scratch, a satisfying common generalization of the theorems of Wintenberger and Tate.

In Corollaries 3.12 and 3.14, we give a couple of applications to lifting λ-adic realizations such that the associated "similitude characters" (e.g., determinant or Clifford norm) of the lifts form strongly compatible systems. Note that even in the case of the classical Kuga–Satake construction, this compatibility is only achieved as a consequence of having an arithmetic descent of the (Hodge-theoretically defined) Kuga–Satake abelian variety; such a descent depends on the deformation theory of K3 surfaces and monodromy arguments (due to Deligne [1972] and André [1996a]).

Our final result is the promised motivic application:

Corollary 1.6. *Let F be a totally imaginary number field, let E be a number field, and let $\mathcal{G}_{F,E}$ denote the motivic Galois group, defined by André's motivated cycles, of pure motives over F with coefficients in E (see Section 2). Let $\widetilde{H} \to H$ be a surjection of linear algebraic groups over E whose kernel is a central torus in $\widetilde{H}$, and let $\rho : \mathcal{G}_{F,E} \to H$ be any motivic Galois representation, with associated λ-adic realizations $\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(\overline{E}_\lambda)$ for some finite set S of places of F. Then there exist a finite, independent of λ, set $P \supset S$ of places of F and, for all λ, lifts*

$$\begin{tikzcd} & \widetilde{H}(\overline{E}_\lambda) \arrow[d] \\ \Gamma_{F,P\cup S_\lambda} \arrow[ur, dashed, "\tilde{\rho}_\lambda"] \arrow[r, "\rho_\lambda"'] & H(\overline{E}_\lambda) \end{tikzcd}$$

such that each $\tilde{\rho}_\lambda$ is de Rham at all places in S_λ, and is moreover crystalline at all places in $S_\lambda \setminus (S_\lambda \cap P)$.

Now suppose $H' \to H$ is a surjection of linear algebraic groups whose kernel is central but of arbitrary multiplicative type, and let $\rho : \mathcal{G}_{F,E} \to H$ again be a motivic Galois representation. Assume that the labeled Hodge cocharacters of ρ (see Definition 2.1) lift to H'. Then there exist a finite set $P \supset S$ of places of F, a finite extension F'/F, and, for all λ, lifts

$$\begin{tikzcd} & H'(\overline{E}_\lambda) \arrow[d] \\ \Gamma_{F',P\cup S_\lambda} \arrow[ur, dashed, "\tilde{\rho}_\lambda"] \arrow[r, "\rho_\lambda|_{\Gamma_{F'}}"'] & H(\overline{E}_\lambda) \end{tikzcd}$$

such that $\tilde{\rho}_\lambda$ is de Rham at all places above S_λ, and is moreover crystalline at all places above $S_\lambda \setminus (S_\lambda \cap P)$.

Remark 1.7. Even admitting a strong finiteness conjecture, that there are finitely many isomorphism classes of ρ (as in Corollary 1.6), having coefficients in E, prescribed Hodge–Tate cocharacters, and good reduction outside a fixed finite set S, Theorem 1.3 still says rather more than Corollary 1.6, since even for fixed λ it applies to infinitely many distinct ρ_λ simultaneously (because we have not bounded the coefficients: recall the example of modular forms of weight two whose nebentypus characters have unbounded conductor, even though supported on the fixed finite set S of primes).

We close this introduction by emphasizing what we do *not* prove. The realizations $\{\rho_\lambda\}_\lambda$ of ρ should, moreover, form a weakly compatible system of Galois representations in the sense that the conjugacy class of $\rho_\lambda(fr_v)$ is defined over $\overline{E}$ and is suitably independent of λ (for v outside $S \cup S_\lambda$), and in turn one would hope to construct lifts $\tilde{\rho}_\lambda$ with the same frobenius compatibility. This problem seems to be out of reach: I know of no way to establish such results using only Galois-theoretic techniques, although indeed they would follow (assuming the standard conjectures) from the generalized Kuga–Satake conjecture. Alternatively, it is possible to establish results of this nature in settings where ρ_λ and $\tilde{\rho}_\lambda$ are constructed as automorphic Galois representations: the most significant example of this is the recent work of Kret and Shin [2016] associating GSpin-value Galois representations to certain (discrete series at infinity, Steinberg at some finite place) cuspidal automorphic representations of $\mathrm{GSp}_{2n}(\boldsymbol{A}_F)$, for F totally real. Crucially, their construction uses the already-known construction of (SO_{2n+1}-valued) Galois representations for the restrictions of such automorphic representations to $\mathrm{Sp}_{2n}(\boldsymbol{A}_F)$, so it is very much in the spirit of the generalized Kuga–Satake lifting problem.

2. Hodge symmetry

In this section we establish a motivic setting in which our general Galois-theoretic results apply; this setting will both serve as motivation for subsequent sections and allow us to deduce Corollary 1.6 from Theorem 1.3 (and Corollary 3.18). The reader who does not find the motivic language illuminating can safely skip this section.

Rather than working with (pure) homological motives and assuming the standard conjectures, we work with a category of motives that is unconditionally semisimple and Tannakian — and in which we can prove unconditional results — but that would, under the standard conjectures, turn out to be equivalent to the category of homological motives. Namely, let $\mathcal{M}_{F,E}$ denote André's category of motivated motives over F with coefficients in E; see [André 1996b]. We begin by elaborating on the consequences of Hodge symmetry in $\mathcal{M}_{F,E}$. Throughout this discussion, it will be convenient to fix embeddings $\tau_0 : F \hookrightarrow \overline{E}$ and $\iota_\infty : \overline{E} \hookrightarrow \mathbb{C}$. The composite

$\iota_\infty\tau_0 : F \hookrightarrow \mathbb{C}$ yields a Betti fiber functor $H_{\iota_\infty\tau_0} : \mathcal{M}_{F,E} \to \mathrm{Vect}_E$, making $\mathcal{M}_{F,E}$ into a neutral Tannakian category over E. We denote by $\mathcal{G} = \mathcal{G}_B(\iota_\infty\tau_0)$ the associated Tannakian group (tensor automorphisms of the fiber functor), so that $H_{\iota_\infty\tau_0}$ induces an equivalence of tensor categories $\mathcal{M}_{F,E} \xrightarrow{\sim} \mathrm{Rep}(\mathcal{G})$.

We will consider other cohomological realizations on $\mathcal{M}_{F,E}$, and their comparisons with the Betti fiber functor. Let $H_{dR} : \mathcal{M}_{F,E} \to \mathrm{Fil}_{F\otimes_{\mathbb{Q}}E}$ denote the de Rham realization, taking values in filtered $F\otimes_{\mathbb{Q}}E$-modules, and for each place λ of E, let H_λ denote the λ-adic realization, which takes values in finite E_λ-modules with a continuous action of Γ_F. For all embeddings $\tau : F \hookrightarrow \overline{E}$, we obtain an $\overline{E}$-valued fiber functor

$$\omega_{dR,\tau} : M \mapsto \mathrm{gr}^\bullet(e_\tau H_{dR}(M)),$$

where e_τ is the idempotent induced by $\tau\otimes 1 : F\otimes_{\mathbb{Q}}E \to \overline{E}$. Let $\mathcal{G}_{dR}(\tau) = \mathrm{Aut}^{\otimes}(\omega_{dR,\tau})$ be the associated Tannakian group over $\overline{E}$. Of course this fiber functor factors through the category $\mathrm{Gr}_{\overline{E}}$ of graded $\overline{E}$-vector spaces

$$\begin{tikzcd} \mathcal{M}_{F,\overline{E}} \arrow[rr] \arrow[dr] & & \mathrm{Gr}_{\overline{E}} \arrow[dl] \\ & \mathrm{Vect}_{\overline{E}} & \end{tikzcd}$$

so we obtain a corresponding homomorphism $\mu_\tau : \boldsymbol{G}_{m,\overline{E}} \to \mathcal{G}_{dR}(\tau)$. Without specifying τ, we obtain a fiber functor (see [Deligne and Milne 1982, §3]) $\omega_{dR} = \mathrm{gr}^\bullet H_{dR}$ on $\mathcal{M}_{F,E}$ valued in projective $F\otimes_{\mathbb{Q}}E$-modules. By [loc. cit., Theorem 3.2], the functor $\underline{\mathrm{Hom}}^{\otimes}(H_{\iota_\infty\tau_0}, \omega_{dR})$ is a $\mathcal{G}$-torsor over $F\otimes_{\mathbb{Q}}E$. In particular, for all $\tau : F \hookrightarrow \overline{E}$, we can choose a point of $\underline{\mathrm{Hom}}^{\otimes}(H_{\iota_\infty\tau_0}\otimes_E \overline{E}, \omega_{dR,\tau})$ to induce a cocharacter μ_τ of $\mathcal{G}_{\overline{E}}$, and the conjugacy class $[\mu_\tau]$ of μ_τ is independent of this choice.

Definition 2.1. For each $\tau : F \hookrightarrow \overline{E}$, we call any $\mu_\tau : \boldsymbol{G}_{m,\overline{E}} \to \mathcal{G}_{\overline{E}}$ as above a τ-labeled Hodge cocharacter; it is a representative of the conjugacy class of cocharacters $[\mu_\tau]$, the latter being canonically independent of any of the above choices of isomorphisms of fiber functors.

Lemma 2.2. *For all $\sigma \in \mathrm{Gal}(\overline{E}/E)$, $[\mu_\tau] = [\mu_{\sigma\tau}]$. In particular, $[\mu_\tau]$ only depends on the restriction of τ to the maximal CM (or totally real) subfield F_{cm} of F.*

Proof. We decompose $F\otimes_{\mathbb{Q}}E = \prod_i E_i$ into a product of fields, writing p_i for the projection onto E_i. Any E-algebra homomorphism $\tau : F\otimes_{\mathbb{Q}}E \to \overline{E}$ factors through $p_{i(\tau)}$ for a unique $i(\tau)$, and then the $\mathrm{Gal}(\overline{E}/E)$-orbit of τ is precisely those E-algebra homomorphisms (i.e., embeddings $F \hookrightarrow \overline{E}$) $\tau' : F\otimes_{\mathbb{Q}}E \to \overline{E}$ such that $i(\tau) = i(\tau')$. The first claim follows, since both $\omega_{dR,\tau}$ and $\omega_{dR,\sigma\tau}$ can be factored through $p_{i(\tau)}\circ\omega_{dR}$. The second claim follows from the first, and the fact that all motives arise by scalar extension from motives with coefficients in CM (or totally real) fields [Patrikis 2016c, Lemma 4.1.22]. $\square$

Next note that the canonical weight-grading on $\mathcal{M}_{F,E}$ induces a *central* weight homomorphism

$$\omega : \boldsymbol{G}_{m,E} \to \mathcal{G},$$

and likewise for any other choice of fiber functor and Tannakian group (because ω is central, it is in fact canonically independent of any choice of isomorphism between fiber functors). Hodge symmetry then results from the complex conjugation action on Betti cohomology, interpreted via the Betti–de Rham comparison isomorphism, which is a distinguished $\mathbb{C}$-point of $\underline{\mathrm{Hom}}^{\otimes}(\omega_{dR,\tau} \otimes_{\overline{E},\iota_\infty} \mathbb{C}, H_{\iota_\infty \tau} \otimes_{E,\iota_\infty} \mathbb{C})$. Namely, complex conjugation on complex-analytic spaces induces (see [Patrikis 2016c, Lemma 4.1.24]) natural isomorphisms (without restricting to particular graded pieces for the weight and Hodge filtrations, these are isomorphisms of fiber functors over $\mathbb{C}$)

$$\mathrm{gr}^p(e_\tau H^w_{dR}(M)) \otimes_{\overline{E},\iota_\infty} \mathbb{C} \xrightarrow{\sim} \mathrm{gr}^{w-p}(e_{c\tau} H^w_{dR}(M) \otimes_{\overline{E},\iota_\infty} \mathbb{C}), \tag{1}$$

where $c \in \mathrm{Aut}(\overline{E})$ is the choice of complex conjugation for which $\overline{\iota_\infty}\tau = \iota_\infty c\tau$. We deduce the following relation:

Lemma 2.3. *For any embedding $\tau : F \hookrightarrow \overline{E}$, and any choice of complex conjugation $c \in \mathrm{Aut}(\overline{E})$, the conjugacy classes of cocharacters $[\mu_\tau]$ and $[\mu_{c\tau}]$ satisfy*

$$[\mu_\tau] = \omega \cdot [\mu_{c\tau}^{-1}],$$

where ω is the weight cocharacter.

Proof. For the choice of complex conjugation specified by $\overline{\iota_\infty}\tau = \iota_\infty c\tau$, the relation $[\mu_\tau] = \omega \cdot [\mu_{c\tau}^{-1}]$ follows, after base extension $\iota_\infty : \overline{E} \to \mathbb{C}$, from (1) above; but this relation necessarily descends to $\overline{E}$, since the conjugacy classes of cocharacters are defined over any algebraically closed subfield of $\mathbb{C}$. It only remains to observe that $[\mu_{c\tau}]$ is independent of the choice of complex conjugation on $\overline{E}$. This follows from the second assertion of Lemma 2.2. $\square$

The comparison isomorphisms of p-adic Hodge theory then imply that the analogue of Lemma 2.3 also holds for the associated Hodge–Tate cocharacters. For any place λ of E, fix an algebraic closure $\overline{E}_\lambda$. Embeddings $\tau : F \hookrightarrow \overline{E}$ and $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$ then induce $\tau_{\iota_\lambda} : F_v \hookrightarrow \overline{E}_\lambda$ for a suitable place v of F of the same residue characteristic p as λ. Meanwhile, the restriction to Γ_{F_v} of the λ-adic realization induces

$$\mathcal{M}_{F,E} \xrightarrow{H_\lambda|_{\Gamma_{F_v}}} \mathrm{Rep}^{dR}_{E_\lambda}(\Gamma_{F_v}) \xrightarrow{D_{dR}} \mathrm{Fil}_{F_v \otimes_{\mathbb{Q}_p} E_\lambda} \xrightarrow{e_{\tau_{\iota_\lambda}}} \mathrm{Fil}_{\overline{E}_\lambda} \xrightarrow{\mathrm{gr}} \mathrm{Gr}_{\overline{E}_\lambda} \longrightarrow \mathrm{Vect}_{\overline{E}_\lambda}, \tag{2}$$

where $D_{dR} : \mathrm{Rep}^{dR}_{E_\lambda}(\Gamma_{F_v}) \to \mathrm{Fil}_{F_v \otimes_{\mathbb{Q}_p} E_\lambda}$ denotes Fontaine's functor restricted to the category of de Rham representations. (Here we have invoked Faltings' p-adic de Rham comparison isomorphism from [Faltings 1989], and the fact — already noted by André [1996b] — that it extends to a comparison isomorphism on all of $\mathcal{M}_{F,E}$;

for details of the latter point, see [Patrikis 2016c, Lemma 4.1.25].) Of course, $\mathrm{Rep}^{dR}_{E_\lambda}(\Gamma_{F_v})$ also has its standard forgetful fiber functor (let us say $\overline{E}_\lambda$-valued), yielding a Tannakian group $\Gamma^{dR}_{v,\lambda}$ for de Rham Γ_{F_v}-representations over $\overline{E}_\lambda$; by choosing an isomorphism between the two $\overline{E}_\lambda$-valued fiber functors on $\mathrm{Rep}^{dR}_{E_\lambda}(\Gamma_{F_v})$, we obtain a canonical conjugacy class (recalling [Deligne and Milne 1982, Theorem 3.2]) of "τ_{ι_λ}-labeled Hodge–Tate cocharacters" $[\mu_{\tau_{\iota_\lambda}}]$ of $\Gamma^{dR}_{v,\lambda}$. Specializing, this construction defines the labeled Hodge–Tate cocharacters of any de Rham Galois representation $\rho : \Gamma_{F_v} \to H(E_\lambda)$, for any affine algebraic group H over E_λ.

To relate the τ_{ι_λ}-labeled Hodge–Tate cocharacters in the motivic setting to the Hodge cocharacters previously discussed, note that the de Rham comparison isomorphism [Faltings 1989] yields a natural isomorphism of tensor functors $\mathcal{M}_{F,E} \to \mathrm{Gr}_{\overline{E}_\lambda}$:

$$\mathrm{gr}(e_\tau(H_{dR}(M) \otimes_E \overline{E}) \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda) \cong \mathrm{gr}(e_{\tau_{\iota_\lambda}}(D_{dR}(H_\lambda(M)|_{\Gamma_{F_v}}) \otimes_{E_\lambda} \overline{E}_\lambda)).$$

We deduce the following corollary.

Corollary 2.4. *For any embedding $\tau : F \hookrightarrow \overline{E}$, and any embedding $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$, there is an equality of conjugacy classes*

$$[\mu_\tau \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda] = [\mu_{\tau_{\iota_\lambda}}].$$

In particular, for all λ, and for all E-embeddings $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$, the conjugacy classes $[\mu_{\tau_{\iota_\lambda}}]$ are independent of (λ, ι_λ) when regarded as valued in the common group $\mathcal{G}_{\overline{E}}$.

3. Lifting

In this section we prove our main results. First we recall the setting and some notational conventions that will be in effect for the rest of the paper.

Let F be a totally imaginary field (see Remark 1.4), let S be a finite set of places of F containing the infinite places, and let E be any number field. Fix an algebraic closure $\overline{F}$, and set $\Gamma_F = \mathrm{Gal}(\overline{F}/F)$. We write $F(S)$ for the maximal extension of F inside $\overline{F}$ that is unramified outside of S, and we set $\Gamma_{F,S} = \mathrm{Gal}(F(S)/F)$. We denote the ring of S-integers in F by $\mathcal{O}_F[1/S]$. We also set $F_S = \prod_{v \in S} F_v$. If L is a finite extension of F (inside $\overline{F}$), we then abusively continue to write S for the set of all places of L above those in S, with corresponding notation $L(S)$, $\Gamma_{L,S}$, etc. For a place λ of E, let S_λ denote the set of places of F with the same residue characteristic as λ. We freely use the terminology established in Definition 1.1.

Torus quotients. In this subsection we prove Theorem 1.3. We begin in the next few paragraphs by gathering together all of the "independent of λ and ρ_λ" data, and the auxiliary constructions we make on top of this data. Continuing with

F, S, and E as above, also fix a surjection $\widetilde{H} \to H$ of linear algebraic groups over E whose kernel is a central torus, which we denote by C. (We will without comment also write C for the base change to various algebraically closed fields containing E.) Next fix an isogeny complement H_1 of C in $\widetilde{H}$ (for existence of such H_1, see [Conrad 2011, Proposition 5.3, Step 1]); thus, $H_1 \cdot C = \widetilde{H}$, and $H_1 \cap C$ is finite. For technical reasons, we will later want to include all primes dividing $\#(H_1 \cap C)(\overline{E})$ in the set of bad primes S; this will be indicated at the necessary point (see the discussion following Lemma 3.5), but it does no harm simply to add these primes to S from now. Consider the quotient map $\xi : \widetilde{H} \to \widetilde{H}/H_1$; $\widetilde{Z}^\vee = \widetilde{H}/H_1$ is a torus, and there is an isogeny $C \to \widetilde{Z}^\vee$, with kernel $C \cap H_1$. Fix a split torus $\widetilde{Z}$ over F whose dual group (constructed over $\overline{E}$) is isomorphic to $\widetilde{Z}^\vee \otimes_E \overline{E}$, and fix such an identification (implicit from now on).

Fix a set of cocharacters $\{\mu_\tau : \boldsymbol{G}_{m,\overline{E}} \to H_{\overline{E}}\}_{\tau : F \hookrightarrow \overline{E}}$, and a central cocharacter $\omega : \boldsymbol{G}_{m,\overline{E}} \to H_{\overline{E}}$, satisfying the Hodge symmetry requirement of item (2) of Definition 1.1. Denote by F_{cm} the maximal CM subfield of F. The condition in Definition 1.1 implies that the cocharacter μ_τ depends only on the restriction of τ to F_{cm}; we denote this restriction by $\tau_{cm} : F_{cm} \hookrightarrow \overline{E}$. We fix a set of representatives I of $\mathrm{Hom}(F_{cm}, \overline{E})$ modulo complex conjugation, and for each $\sigma \in I$, we fix a lift $\widetilde{\mu_\sigma}$ to $\widetilde{H}$ of μ_σ, as well as a (central) lift $\widetilde{\omega}$ of ω. Note that this is possible, because C is a torus. If $\tau : F \hookrightarrow \overline{E}$ restricts to a $\sigma \in I$, we then set $\widetilde{\mu_\tau} = \widetilde{\mu_\sigma}$; if not, then $c\tau : F \hookrightarrow \overline{E}$ restricts to a $\sigma \in I$, and we then set $\widetilde{\mu_\tau} = \widetilde{\omega}\widetilde{\mu_\sigma}^{-1}$.

Lemma 3.1. *Fix once and for all an embedding $\iota_\infty : \overline{E} \hookrightarrow \mathbb{C}$. There exists an algebraic automorphic representation ψ of $\widetilde{Z}(\boldsymbol{A}_F)$ such that for all $\tau : F \hookrightarrow \overline{E}$, inducing $\tau_{\iota_\infty} : F_v \hookrightarrow \mathbb{C}$ by composition with ι_∞, the local component $\psi_v : F_v^\times \to \mathbb{C}^\times$ is given by*

$$\psi_v(z) = \tau_{\iota_\infty}(z)^{\xi(\widetilde{\mu_\tau})}\overline{\tau_{\iota_\infty}}(z)^{\xi(\widetilde{\mu_{c\tau}})}.$$

(*Recall that ξ is the quotient $\widetilde{H} \to \widetilde{Z}^\vee$.*)

Proof. We readily reduce to the case $\widetilde{Z} = \boldsymbol{G}_m$, where it follows from the description, due to Weil [1956], of the possible archimedean components of algebraic Hecke characters. (This is where Hodge-symmetry is required.) □

From now on we fix such a ψ, and we let T denote the finite set of places of F such that ψ is unramified outside T. For any embedding $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$, we can then consider the λ-adic realization[3]

$$\psi_{\iota_\lambda} : \Gamma_{F,T \cup S_\lambda} \to \widetilde{Z}^\vee(\overline{E}_\lambda).$$

Each ψ_{ι_λ} is a geometric Galois representation, with good reduction outside T, and for any $\tau : F \hookrightarrow \overline{E}$, inducing $\tau_{\iota_\lambda} : F_v \hookrightarrow \overline{E}_\lambda$, the Hodge–Tate cocharacter of ψ_{ι_λ} associated to τ_{ι_λ} is $\xi(\widetilde{\mu_\tau}) \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda$.

[3]To be precise, this depends on ι_∞ and ι_λ; but ι_∞ is fixed throughout the paper.

Now we consider any geometric representation

$$\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(\bar{E}_\lambda)$$

having good reduction outside S, along with an embedding $\iota_\lambda : \bar{E} \hookrightarrow \bar{E}_\lambda$ such that the Hodge–Tate cocharacters of ρ_λ arise from the collection $\{\mu_\tau : \boldsymbol{G}_{m,\bar{E}} \to H_{\bar{E}}\}_{\tau: F\hookrightarrow \bar{E}}$ via ι_λ. Because the kernel of $\widetilde{H} \to H$ is a central torus, a fundamental theorem of Tate (see [Serre 1977, §6]) ensures in this case that ρ_λ, *as a representation of* Γ_F, lifts to $\widetilde{H}$. As we will see, our arguments in fact imply Tate's theorem (Corollary 3.9), so we do not need to assume it in what follows.

We can define an obstruction class $\mathcal{O}(\rho_\lambda)$ to lifting ρ_λ to a continuous representation $\Gamma_{F,S\cup S_\lambda} \to H_1(\bar{E}_\lambda)$ in the usual way: choose a topological (but not group-theoretic) lift ρ'_λ, and then form the 2-cocycle $(g,h) \mapsto \rho'_\lambda(gh)\rho'_\lambda(h)^{-1}\rho'_\lambda(g)^{-1}$, defining

$$\mathcal{O}(\rho_\lambda) \in H^2(\Gamma_{F,S\cup S_\lambda}, H_1 \cap C).$$

Here and in what follows, we simply write $H_1 \cap C$ for the $\bar{E}_\lambda$-points of this finite group scheme.

Remark 3.2. Here lies the essential difficulty to be overcome: while Tate's theorem allows us to annihilate the cohomology classes $\mathcal{O}(\rho_\lambda)$ — after allowing some additional ramification and enlarging the subgroup $H_1 \cap C$ of C — we have to carry out this annihilation in a way that is independent of λ, and moreover, for fixed λ independent of ρ_λ. Simultaneous annihilation of the $\mathcal{O}(\rho_\lambda)$ using only a uniform, finite enlargement of the allowable ramification set and of the subgroup of C in fact does not seem to be possible; we will as a first step have to define modified versions of these obstruction classes that take into account the Hodge numbers of ψ.

Before proceeding, we reinterpret the obstruction $\mathcal{O}(\rho_\lambda)$ (we will only use the local version of what follows; in particular, the arguments of the present section depend only on the local version of Tate's theorem, which is an almost immediate consequence of local duality).

Lemma 3.3. *Let v be a finite place of F, and suppose that $\tilde{\rho}_\lambda : \Gamma_{F_v} \to \widetilde{H}(\bar{E}_\lambda)$ is any continuous homomorphism lifting $\rho_\lambda|_{\Gamma_{F_v}}$. Then $\mathcal{O}(\rho_\lambda)|_{\Gamma_{F_v}}$ is equal to the inverse of $\mathcal{O}(\xi(\tilde{\rho}_\lambda))$, the obstruction associated to lifting $\xi(\tilde{\rho}_\lambda) : \Gamma_{F_v} \to \widetilde{Z}^\vee(\bar{E}_\lambda)$ to C.* (*The same holds if we replace Γ_{F_v} by Γ_F, but we do not require this.*)

Proof. Before beginning the proof proper, we make precise our convention for coboundary maps: the *inverse* appearing in the conclusion of the lemma is crucial, and it is easy to get confused if one is not careful with the definitions. Let Γ be a group and M a (for simplicity) trivial Γ-module. For a function $\alpha : \Gamma^n \to M$, set

$$\delta(\alpha)(g_1, \dots, g_{n+1}) = \alpha(g_2, \dots, g_n) + \sum_{i=1}^{n} (-1)^i \alpha(g_1, \dots, g_i g_{i+1}, \dots, g_{n+1}) + (-1)^{n+1}\alpha(g_1, \dots, g_n).$$

For $n=1$, this says $\delta(\alpha)(g,h)=\alpha(h)\alpha(gh)^{-1}\alpha(g)$, and in the situation considered above $(g,h)\mapsto \rho'_\lambda(gh)\rho'_\lambda(h)^{-1}\rho'_\lambda(g)^{-1}$ is in fact a 2-cocycle.

Tate's theorem implies that, for sufficiently large m, the image of $\mathcal{O}(\rho_\lambda)$ in $H^2(\Gamma_{F_v}, C[m])$ vanishes, i.e., $\mathcal{O}(\rho_\lambda)=\delta(\phi)$ for some $\phi:\Gamma_{F_v}\to C[m]$. The product $\rho'_\lambda\cdot\phi$ is then a homomorphism $\Gamma_{F_v}\to(H_1\cdot C[m])(\overline{E}_\lambda)$ lifting ρ_λ; we set $\tilde{\rho}_\lambda=\rho'_\lambda\cdot\phi$. Clearly $\xi(\tilde{\rho}_\lambda)=\xi(\phi)$, and then $\mathcal{O}(\xi(\tilde{\rho}_\lambda))$ is (tautologically) represented by the cocycle $(g,h)\mapsto\phi(gh)\phi(h)^{-1}\phi(g)^{-1}$, i.e., by $\delta(\phi)^{-1}=\mathcal{O}(\rho_\lambda)^{-1}\in Z^2(\Gamma_F, H_1\cap C)$.[4] This proves the claim for our particular lift $\tilde{\rho}_\lambda$, but any other lift $\tilde{\rho}^1_\lambda$ gives rise to the same obstruction $\mathcal{O}(\xi(\tilde{\rho}^1_\lambda))$. (The global claim holds for the same reasons, if we admit the global version of Tate's theorem.) □

To address the difficulty indicated in Remark 3.2, we begin by using the abelian representations coming from ψ to construct a second obstruction class. Namely, consider the realization ψ_{ι_λ}, which, for notational simplicity, from now on we simply denote by ψ_λ. The automorphic representation ψ is unramified outside the finite set of places T of F, so ψ_λ is a geometric representation $\Gamma_{F,T\cup S_\lambda}\to\widetilde{Z}^\vee(\overline{E}_\lambda)$, which has good reduction outside T (i.e., is crystalline at primes of S_λ not in T). Via the isogeny $C\to\widetilde{Z}^\vee$, we can then form a cohomology class measuring the obstruction to lifting ψ_λ to C: let ψ'_λ denote a topological lift $\Gamma_{F,T\cup S_\lambda}\to C(\overline{E}_\lambda)$, defining as before a cohomology class

$$\mathcal{O}(\psi_\lambda)\in H^2(\Gamma_{F,T\cup S_\lambda}, H_1\cap C).$$

We can in turn define (via inflation) a cohomology class

$$\mathcal{O}(\rho_\lambda,\psi_\lambda)=\mathcal{O}(\rho_\lambda)\cdot\mathcal{O}(\psi_\lambda)\in H^2(\Gamma_{F,S\cup T\cup S_\lambda}, H_1\cap C),$$

which is represented by the 2-cocycle (recall that C is central in $\widetilde{H}$)

$$(g,h)\mapsto(\rho'_\lambda\cdot\psi'_\lambda)(gh)(\rho'_\lambda\cdot\psi'_\lambda)(h)^{-1}(\rho'_\lambda\cdot\psi'_\lambda)(g)^{-1}.$$

(Note, however, that the function $g\mapsto(\rho'_\lambda\cdot\psi'_\lambda)(g)$ is valued in $\widetilde{H}$, not in H_1.)

We need one more lemma before getting to the crucial local result (Lemma 3.5).

Lemma 3.4. *For all places $v\in S_\lambda$, and for any choice of embedding $\iota_\lambda:\overline{E}\hookrightarrow\overline{E}_\lambda$, there exists a de Rham lift*

$$\begin{tikzcd} & \widetilde{H}(\overline{E}_\lambda) \arrow[d] \\ \Gamma_{F_v} \arrow[ur, dashed, "\tilde{\rho}_\lambda"] \arrow[r, "\rho_\lambda"'] & H(\overline{E}_\lambda) \end{tikzcd}$$

of $\rho_\lambda|_{\Gamma_{F_v}}$ such that for all embeddings $\tau_\lambda:F_v\hookrightarrow\overline{E}_\lambda$, the τ_λ-labeled Hodge–Tate

[4]Note that ϕ is valued in $C[m]$, not $H_1\cap C$, so $\delta(\phi)$ need not be a coboundary in $Z^2(\Gamma_{F_v}, H_1\cap C)$.

cocharacter of $\tilde{\rho}_\lambda$ is (conjugate to) $\widetilde{\mu_\tau} \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda$, where $\tau : F \hookrightarrow \overline{E}$ is defined by the diagram

$$\begin{tikzcd} F_v \arrow[r, "\tau_\lambda"] & \overline{E}_\lambda \\ F \arrow[u] \arrow[r, "\tau"] & \overline{E} \arrow[u, "\iota_\lambda"'] \end{tikzcd}$$

Moreover, if ρ_λ is crystalline, then $\tilde{\rho}_\lambda$ may be taken to be crystalline.

Proof. For each $\tau_\lambda : F_v \hookrightarrow \overline{E}_\lambda$, set for notational simplicity $\widetilde{\mu_{\tau_\lambda}} = \widetilde{\mu_\tau} \otimes_{\overline{E},\iota_\lambda} \overline{E}_\lambda$, where τ is determined as in the diagram, and where $\widetilde{\mu_\tau}$ is the lift of μ_τ we have fixed above. The proof of [Patrikis 2016c, Corollary 3.2.12] shows that for any collection of cocharacters lifting the Hodge cocharacters of ρ_λ, and in particular for our $\widetilde{\mu_{\tau_\lambda}}$, there exists a Hodge–Tate lift $\tilde{\rho}_\lambda : \Gamma_{F_v} \to \widetilde{H}(\overline{E}_\lambda)$ whose τ_λ-labeled Hodge–Tate cocharacter is $\widetilde{\mu_{\tau_\lambda}}$. Now consider the isogeny lifting problem

$$\begin{tikzcd} & \widetilde{H}(\overline{E}_\lambda) \arrow[d] \\ \Gamma_{F_v} \arrow[ur, dashed] \arrow[r, "(\rho_\lambda,\xi(\tilde{\rho}_\lambda))"'] & H(\overline{E}_\lambda) \times \widetilde{Z}^\vee(\overline{E}_\lambda) \end{tikzcd}$$

Since $(\rho_\lambda, \xi(\tilde{\rho}_\lambda))$ admits a Hodge–Tate lift (namely, $\tilde{\rho}_\lambda$), and is itself de Rham (ρ_λ is de Rham by assumption, and any abelian Hodge–Tate representation is de Rham), we can apply [Conrad 2011, Corollary 6.7] to deduce the existence of a de Rham lift $\tilde{\rho}'_\lambda$, which clearly has the same Hodge–Tate cocharacters as $\tilde{\rho}_\lambda$, since they differ by a finite-order twist. If we further assume ρ_λ is crystalline, then we need only a minor modification to this argument: some power $\xi(\tilde{\rho}_\lambda)^d$ is crystalline, so if we instead consider the problem of lifting the crystalline representation $(\rho_\lambda, [d]\xi(\tilde{\rho}_\lambda))$ through the composite isogeny

$$\widetilde{H} \to H \times \widetilde{Z}^\vee \xrightarrow{\mathrm{id}\times[d]} H \times \widetilde{Z}^\vee,$$

then again [Conrad 2011, Corollary 6.7] applies to produce a crystalline lift of ρ_λ with the desired Hodge–Tate cocharacters. $\square$

Here is the key lemma:

Lemma 3.5. *For any place $v \in S_\lambda$ not belonging to the finite set $S \cup T$, the restriction $\mathcal{O}(\rho_\lambda, \psi_\lambda)|_{\Gamma_{F_v}}$ is trivial.*

Proof. Under the assumption on v, both ρ_λ and ψ_λ are crystalline at v. Lemma 3.4 above shows that $\rho_\lambda|_{\Gamma_{F_v}}$ admits a crystalline lift $\tilde{\rho}_\lambda : \Gamma_{F_v} \to \widetilde{H}(\overline{E}_\lambda)$ such that $\xi(\tilde{\rho}_\lambda)$ has the same (labeled) Hodge–Tate cocharacters as $\psi_\lambda|_{\Gamma_{F_v}}$. Since they are both crystalline, it follows (see [Chai et al. 2014, 3.9.7 Corollary]) that $\xi(\tilde{\rho}_\lambda) \cdot \psi_\lambda^{-1}|_{\Gamma_{F_v}}$

is unramified; this is an elaboration of the familiar fact that a crystalline character whose Hodge–Tate weights are zero must be unramified. In particular, replacing the initial lift $\tilde{\rho}_\lambda|_{\Gamma_{F_v}}$ by an unramified twist, we may assume $\xi(\tilde{\rho}_\lambda) = \psi_\lambda$ as homomorphisms $\Gamma_{F_v} \to \widetilde{Z}^\vee(\bar{E}_\lambda)$. But recall that Lemma 3.3 implies that $\mathcal{O}(\rho_\lambda) = \mathcal{O}(\xi(\tilde{\rho}_\lambda))^{-1}$, so we deduce that $\mathcal{O}(\rho_\lambda) \cdot \mathcal{O}(\psi_\lambda)|_{\Gamma_{F_v}}$ is trivial. □

Since the set of places $S \cup T$ is finite, by the local version of Tate's theorem, the vanishing of $H^2(\Gamma_{F_v}, \mathbb{Q}/\mathbb{Z})$ for all places v of F, we may enlarge $H_1 \cup C$ to some $C[m]$ inside the torus C so as to kill the image of $H^2(\Gamma_{F_v}, H_1 \cap C) \to H^2(\Gamma_{F_v}, C[m])$ for all $v \in S \cup T$. (We emphasize that m only depends on the set of places $S \cup T$ of F and the finite group $H_1 \cap C$.) It follows then from Lemma 3.5 that if λ does not belong to $S \cup T$, then $\mathcal{O}(\rho_\lambda, \psi_\lambda)$ in fact belongs to

$$Ш^2_{S \cup T \cup S_\lambda}(F, C[m]) = \ker\left(H^2(\Gamma_{F, S \cup T \cup S_\lambda}, C[m]) \to \bigoplus_{v \in S \cup T \cup S_\lambda} H^2(\Gamma_{F_v}, C[m])\right).$$

We can moreover guarantee that this holds regardless of λ by an additional finite enlargement of m (since the number of exceptional λ is finite). Furthermore, by including the primes dividing $\#(H_1 \cap C)$ in $S \cup T$ (if necessary), we can assume that m is divisible only by primes in $S \cup T$. (Note that inflation to allow additional primes of ramification still has image in the corresponding Shafarevich–Tate group, since Γ_{F_v}/I_{F_v} has cohomological dimension one for all finite places v.) Thus, after these uniform enlargements of m and $S \cup T$ (which we do not reflect in the notation), we have $\mathcal{O}(\rho_\lambda, \psi_\lambda) \in Ш^2_{S \cup T \cup S_\lambda}(F, C[m])$.

We are now in a position to apply global duality to analyze the cohomology group $Ш^2_{S \cup T \cup S_\lambda}(F, C[m])$. We will need, however, to allow still more primes of ramification in order to kill the class $\mathcal{O}(\rho_\lambda, \psi_\lambda)$; the following crucial lemma allows us to do this in a way that does not depend on λ, but before stating the lemma, we have to recall the Grunwald–Wang theorem (in a somewhat specialized form).

Theorem 3.6 (Grunwald–Wang; see Theorem X.1 of [Artin and Tate 1968]). *Let F be a number field, and let m be a positive integer. Then an element $x \in F^\times$ belongs to $(F^\times)^m$ if and only if x is in $(F_v^\times)^m$ for all places v of F, except when all three of the following conditions, referred to as the special case, hold for the pair (F, m):*

- *Let s_F denote the largest integer r such that $\eta_r = \zeta_{2^r} + \zeta_{2^r}^{-1}$ is an element of F (here ζ_{2^r} denotes a primitive 2^r-th root of unity). Then -1, $2 + \eta_{s_F}$, and $-(2 + \eta_{s_F})$ are nonsquares in F.*
- $\mathrm{ord}_2(m) > s_F$.
- *The set of 2-adic places of F at which -1, $2 + \eta_{s_F}$, and $-(2 + \eta_{s_F})$ are nonsquares in F is empty.*

In the special case, the element $(2+\eta_{s_F})^{m/2}$ *is the unique (up to* $(F^\times)^m$*-multiple) counterexample to the local-global principle for* m*-th powers in* $F^\times$.

Here is the lemma:

Lemma 3.7. *Recall that* $S \cup T$ *is a fixed finite set of places of the number field* F*, and that* m *is a fixed integer. Let* V *be a finite set of finite places of* F *such that*

- *all elements of* V *are unramified in* $F(\mu_m)$,
- *the places of* $F(\mu_m)$ *lying above* V *generate the class group of* $F(\mu_m)$*, and*
- *every element of* $\mathrm{Gal}(F(\mu_m)/F)$ *is equal to a (geometric, say) frobenius element at* v *for some* $v \in V$.

Then for all places λ *of* E *we can deduce:*

(1) *If* (F, m) *is not in the Grunwald–Wang special case,* $Ш^2_{S\cup T\cup V\cup S_\lambda}(F, C[m])$ *is trivial.*

(2) *If* (F, m) *is in the Grunwald–Wang special case, then the image of the canonical map*

$$Ш^2_{S\cup T\cup V\cup S_\lambda}(F, C[m]) \to Ш^2_{S\cup T\cup V\cup S_\lambda}(F, C[2m])$$

is trivial.

Proof. First note that such sets V exist, by finiteness of the class number and the Čebotarev density theorem. Since (all places of F above) the primes dividing m are contained in $S \cup T$, an application of Poitou–Tate duality immediately reduces us to showing (as a Galois module, $C[m]$ is $\dim(C)$ copies of $\mathbb{Z}/m$) the following cases:

(1) If (F, m) is not in the Grunwald–Wang special case, then

$$Ш^1_{S\cup T\cup V\cup S_\lambda}(F, \mu_m) = 0.$$

(2) If (F, m) is in the special case, then the map

$$Ш^1_{S\cup T\cup V\cup S_\lambda}(F, \mu_{2m}) \to Ш^1_{S\cup T\cup V\cup S_\lambda}(F, \mu_m)$$

induced by $\mu_{2m} \xrightarrow{2} \mu_m$ is zero.

We first restrict to $\Gamma_{F(\mu_m), S\cup T\cup V\cup S_\lambda}$ (note that this is actually restriction to a subgroup, since $F(\mu_m)/F$ is ramified only at primes in $S \cup T$), obtaining an element of $Ш^1_{S\cup T\cup V\cup S_\lambda}(F(\mu_m), \mu_m)$. After this restriction, as we will see, the Grunwald–Wang theorem does not intervene.

To lighten the notation in the rest of the proof, we define $L = F(\mu_m)$ and $Q_\lambda = S \cup T \cup V \cup S_\lambda$. We also refer the reader to the notation established at the beginning of Section 3. Recall that $F(Q_\lambda)$ denotes the maximal extension of F inside $\overline{F}$ that is unramified outside Q_λ; it contains L. Let $\mathcal{O}_{F(Q_\lambda)}$ denote the ring of Q_λ-integers in $F(Q_\lambda)$ (i.e., the elements of $F(Q_\lambda)$ that are integral outside of places above Q_λ). We then have an exact (Kummer theory) sequence

$$1 \to \mu_m \to \mathcal{O}_{F(Q_\lambda)}^\times \xrightarrow{m} \mathcal{O}_{F(Q_\lambda)}^\times \to 1,$$

and the corresponding long exact sequence in Γ_{L,Q_λ}-cohomology yields an isomorphism

$$\mathcal{O}_L\Big[\frac{1}{Q_\lambda}\Big]^\times \Big/ \Big(\mathcal{O}_L\Big[\frac{1}{Q_\lambda}\Big]^\times\Big)^m \xrightarrow{\sim} H^1(\Gamma_{L,Q_\lambda}, \mu_m);$$

critically, surjectivity here follows from the vanishing of $H^1(\Gamma_{L,Q_\lambda}, \mathcal{O}_{F(Q_\lambda)}^\times)$, which itself is a consequence of the natural isomorphism $\mathrm{Cl}_{Q_\lambda}(L) \cong H^1(\Gamma_{L,Q_\lambda}, \mathcal{O}_{F(Q_\lambda)}^\times)$ [Neukirch et al. 2000, Proposition 8.3.11(ii)] and our assumption that V (and hence Q_λ) generates the class group of L. Restricting the Kummer theory isomorphism to classes that are locally trivial at each place of Q_λ, we also obtain the isomorphism

$$\Big(\mathcal{O}_L\Big[\frac{1}{Q_\lambda}\Big]^\times \cap (L_{Q_\lambda}^\times)^m\Big) \Big/ \Big(\mathcal{O}_L\Big[\frac{1}{Q_\lambda}\Big]^\times\Big)^m \xrightarrow{\sim} Ш^1_{Q_\lambda}(L, \mu_m).$$

We claim these groups are trivial. Indeed, take $\alpha \in \mathcal{O}_L[1/Q_\lambda]^\times \cap (L_{Q_\lambda}^\times)^m$, and consider the (abelian) extension $L(\alpha^{1/m})/L$. Global class field theory yields the reciprocity isomorphism

$$\mathbb{A}_L^\times / \big(L^\times N_{L(\alpha^{1/m})/L}(\mathbb{A}_{L(\alpha^{1/m})}^\times)\big) \xrightarrow{\sim} \mathrm{Gal}(L(\alpha^{1/m})/L),$$

but by assumption the source of this map admits a surjection

$$\mathbb{A}_L^\times \Big/ \Big(L^\times L_\infty^\times L_{Q_\lambda}^\times \prod_{w \notin Q_\lambda} \mathcal{O}_{L_w}^\times\Big) \twoheadrightarrow \mathbb{A}_L^\times / \big(L^\times N_{L(\alpha^{1/m})/L}(\mathbb{A}_{L(\alpha^{1/m})}^\times)\big).$$

(At unramified places, the image of the norm map contains the local units; and at places in Q_λ, $L(\alpha^{1/m})/L$ is split.) By assumption ($\mathrm{Cl}_{Q_\lambda}(L) = 0$), the source of this surjection is trivial, so $L(\alpha^{1/m}) = L$, and we deduce that $Ш^1_{Q_\lambda}(L, \mu_m) = 0$.

It follows that inflation identifies the group $Ш^1_{Q_\lambda}(F, \mu_m)$ with the classes in $H^1(\mathrm{Gal}(L/F), \mu_m)$ that are trivial upon restriction to Q_λ. Since every element of $\mathrm{Gal}(L/F)$ is a frobenius element at some prime in $V \subset Q_\lambda$, $Ш^1_{Q_\lambda}(F, \mu_m)$ is actually equal to the set of *everywhere* locally trivial classes

$$Ш^1_{|F|}(F, \mu_m) := \ker\Big(H^1(\Gamma_F, \mu_m) \to \prod_{v \in |F|} H^1(\Gamma_{F_v}, \mu_m)\Big),$$

where $|F|$ denotes the set of all places of F. This is precisely the subject of the Grunwald–Wang theorem, and it is zero if (F, m) is not in the special case. Thus, we need only consider the possibility that (F, m) is in the special case, where $Ш^1_{|F|}(F, \mu_m)$ has order two, and a representative of the nontrivial class is the (image under the Kummer map of the) element $(2 + \eta_{s_F})^{m/2}$ of $(F^\times)^{m/2}$. This description holds regardless of m, so in particular the nontrivial class of $Ш^1_{|F|}(F, \mu_{2m})$ is represented by $(2+\eta_{s_F})^m$. Its image under $\mu_{2m} \xrightarrow{2} \mu_m$, which via Kummer theory is induced by the identity map $F^\times \to F^\times$, is again $(2+\eta_{s_F})^m$, which is now visibly an m-th power, completing the proof.[5] □

We summarize our conclusion, noting that the value of m in the following corollary may be $2m$ in the earlier notation:

Corollary 3.8. *There is an integer m and a finite set of places $Q \supset S \cup T$, both independent of λ and of the choice of ρ_λ having good reduction outside S and the prescribed Hodge–Tate cocharacters $\{\mu_\tau\}_{\tau: F \hookrightarrow \bar{E}}$, such that the image of $\mathcal{O}(\rho_\lambda, \psi_\lambda)$ in $H^2(\Gamma_{F, Q \cup S_\lambda}, C[m])$ is zero.*

Before proceeding, it is worth noting that the argument just given yields a novel proof of the global version of Tate's vanishing theorem (taking as input the much easier local theorem); it is also a stronger proof, yielding an explicit upper bound on how much ramification has to be allowed, and how much the coefficients need to be enlarged, in order to kill a cohomology class in $H^2(\Gamma_{F,V}, \mathbb{Z}/N)$ for some finite set of places V and integer N.

Corollary 3.9. *Let V be a finite set of places of F, and let N be an integer. Then the image of $H^2(\Gamma_{F,V}, \mathbb{Z}/N)$ in $H^2(\Gamma_{F,V \cup W}, \mathbb{Z}/2NM)$ is trivial, where*

- *M is large enough that for all $v \in V$, the image of*
$$H^2(\Gamma_{F_v}, \mathbb{Z}/N) \to H^2(\Gamma_{F_v}, \mathbb{Z}/NM)$$
is zero,[6] *and*
- *once M is fixed as above, W is large enough that*
  - *$V \cup W$ contains (all places above) $2NM$,*
  - *$\mathrm{Cl}_{V \cup W}(F(\mu_{NM})) = 0$, and*
  - *each element of $\mathrm{Gal}(F(\mu_{NM})/F)$ is equal to a frobenius element at w for some $w \in W$.*

(The factors of two are only necessary in the Grunwald–Wang special case.) In particular, $H^2(\Gamma_F, \mathbb{Q}/\mathbb{Z}) = 0$.

Remark 3.10. A different proof of Tate's theorem (without arithmetic duality theorems, but instead relying on a finer study of Hecke characters of F) is given in [Serre 1977, §6.5]. There Serre remarks that Tate originally proved the vanishing theorem using global duality, but further assuming Leopoldt's conjecture; we have of course circumvented Leopoldt here.

[5] In concrete terms, this says that if an element of $F^\times$ is everywhere locally a $(2m)$-th power, then it is globally an m-th power.

[6] This is easy to make explicit, using local duality, in terms of $\mu_\infty(F_v)$.

Now we return to the conclusion of Corollary 3.8. Let $b_\lambda : \Gamma_{F,Q\cup S_\lambda} \to C[m]$ be a cochain trivializing $\mathcal{O}(\rho_\lambda, \psi_\lambda)$. Then

$$\tilde{\rho}_\lambda = \rho'_\lambda \cdot \psi'_\lambda \cdot b_\lambda : \Gamma_{F,Q\cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda)$$

is a homomorphism lifting ρ_λ. We claim that $\tilde{\rho}_\lambda$ is, moreover, de Rham at all places in S_λ. To see this, note that under the isogeny $\widetilde{H} \to H \times \widetilde{Z}^\vee$, $\tilde{\rho}_\lambda$ pushes forward to $(\rho_\lambda, \psi_\lambda \xi(b_\lambda))$, the second coordinate being a finite-order twist of ψ_λ (and in particular, de Rham). But now we can invoke the local results of Wintenberger [1995, §1] and Conrad [2011, Theorem 6.2], asserting that a lift of a de Rham representation through an isogeny is de Rham if and only if the Hodge–Tate cocharacter lifts through the isogeny (which is obviously the case here, as ψ was constructed to ensure this).

Finally, we can refine this to the statement that ρ_λ admits a geometric lift that is moreover crystalline at all places of S_λ, provided S_λ does not intersect a certain finite set of primes that is independent of λ and ρ_λ (but somewhat larger than the set Q we have thus far constructed). This will complete the proof of Theorem 1.3.

Proof of Theorem 1.3. We resume the above discussion. So far we have a constructed geometric lifts $\tilde{\rho}_\lambda : \Gamma_{F,Q\cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda)$, where Q contains $S \cup T$ and whatever other additional primes are needed for the conclusion of Corollary 3.8. The only remaining task is to show that for some (independent of ρ_λ) set P, we can modify the initial lift (by a finite-order twist) to guarantee that it has good reduction outside P. Under the isogeny $\widetilde{H} \to H \times \widetilde{Z}^\vee/\xi(C[m])$, $\tilde{\rho}_\lambda$ pushes forward to

$$\tau_\lambda := (\rho_\lambda, \psi_\lambda \mod \xi(C[m])),$$

which is crystalline for all v in S_λ but not in $S \cup T$. For all $v \in S_\lambda \setminus (S_\lambda \cap (S \cup T))$, $\tilde{\rho}_\lambda|_{\Gamma_{F_v}}$ is of course a de Rham lift of τ_λ, so [Conrad 2011, Theorem 6.2 and Corollary 6.7] (building on [Wintenberger 1995]) shows that $\tau_\lambda|_{\Gamma_{F_v}}$ admits *some* crystalline lift $\tilde{\tau}_{\lambda,v} : \Gamma_{F_v} \to \widetilde{H}(\overline{E}_\lambda)$, and therefore there are finite-order characters $\chi_{\lambda,v} : \Gamma_{F_v} \to C[m]$ such that each $\tilde{\rho}_\lambda|_{\Gamma_{F_v}} \cdot \chi_{\lambda,v}$ is crystalline. We wish to glue the inertial restrictions $\chi_{\lambda,v}|_{I_{F_v}}$ together into a global character, with an independent-of-λ control on the ramification. The cokernel of the restriction map

$$\mathrm{Hom}(\Gamma_{F,Q\cup S_\lambda}, C[m]) \to \bigoplus_{v\in S_\lambda} \mathrm{Hom}(I_{F_v}, C[m])^{\Gamma_{F_v}/I_{F_v}} \tag{3}$$

may be nontrivial;[7] but we will show that any element of the cokernel is annihilated by appropriate enlargements of Q and m.

[7]Of course, we only need to consider the cokernel of the map to the direct sum over $v \in S_\lambda \setminus (S_\lambda \cap (S \cup T))$; to lighten the notation we will work with all $v \in S_\lambda$, taking some arbitrary (e.g., trivial) choice of $\chi_{\lambda,v}$ at any places in $(S \cup T) \cap S_\lambda$.

By the congruence subgroup property for GL_1 (a theorem of Chevalley [1951]), there is an ideal $\mathfrak{n}$ of $\mathcal{O}_F$ such that

$$\{x \in \mathcal{O}_F^\times : x \equiv 1 \pmod{\mathfrak{n}}\} \subseteq (\mathcal{O}_F^\times)^m.$$

Let R be the set of primes supporting $\mathfrak{n}$ (note that $\mathfrak{n}$ and R are independent of ρ_λ!), and set

$$U_R = \left\{(x_v)_{v \in R} \in \prod_{v \in R} \mathcal{O}_{F_v}^\times : x_v \equiv 1 \pmod{\mathfrak{n}} \text{ for all } v \in R\right\}.$$

Then whenever $S_\lambda \cap R = \varnothing$ (so, excluding a finite number of bad λ), consider the character (here and in what follows, we suppress the class field theory identifications)

$$(\chi_{\lambda,v})_{v \in S_\lambda} \times 1 \times 1 \times 1 : \prod_{v \in S_\lambda} \mathcal{O}_{F_v}^\times \times \prod_{v \in R} U_R \times \prod_{v \notin R \cup S_\lambda} \mathcal{O}_{F_v}^\times \times F_\infty^\times \to C[m],$$

which extends by 1 to a character

$$\left(\prod_{v \in S_\lambda} \mathcal{O}_{F_v}^\times \times \prod_{v \in R} U_R \times \prod_{v \notin R \cup S_\lambda} \mathcal{O}_{F_v}^\times \times F_\infty^\times\right) \cdot F^\times \to C[m]$$

(an element of the intersection is a global unit congruent to 1 modulo $\mathfrak{n}$, hence is contained in $(\mathcal{O}_F^\times)^m$, where $\chi_{\lambda,v}$ is obviously trivial). We can then extend from this finite-index subgroup of $\mathbb{A}_F^\times$ to a character $\chi_\lambda : \mathbb{A}_F^\times / F^\times \to \mu_\infty(C)$. In fact, we see that χ_λ can be chosen to be valued in $C[M]$ for m sufficiently large but independent of λ: m can be quantified in terms of the generalized class group of level U_R, but the details do not concern us.

Replacing $\tilde{\rho}_\lambda$ by its finite-order twist

$$\tilde{\rho}_\lambda \cdot \chi_\lambda : \Gamma_{F, Q \cup R \cup S_\lambda} \to \widetilde{H}(\overline{E}_\lambda),$$

we have achieved geometric lifts of ρ_λ with compatible Hodge–Tate cocharacters, and which are crystalline at all places in S_λ outside of $R \cup S \cup T$. □

Remark 3.11. Contrast the final step [Wintenberger 1995, §2.3.5] of Wintenberger's main theorem, where to ensure crystallinity of the lifts he makes a further finite base change on F (having already made several such in order to show lifts exist, as is necessary in his isogeny setup), adding appropriate roots of unity and then passing to a Hilbert class field to kill a cokernel analogous to that of (3). As elsewhere, our argument is orthogonal to Wintenberger's, in allowing additional ramification and larger coefficients, rather than passing to a finite extension of F.

We now deduce some corollaries on finding lifts of ramification-compatible systems whose "similitude characters" (determinant, Clifford norm, etc.) form strongly compatible systems, in the sense that at all finite places their associated Weil group representations are isomorphic (see, e.g., [Barnet-Lamb et al. 2014,

§5.1], where these are called *strictly* compatible). As with Theorem 1.3 and the preceding results, we show a somewhat stronger finiteness result, which applies to all representations with good reduction outside a fixed finite set S. These corollaries will follow from the above results and the Hermite–Minkowski finiteness theorem.

Corollary 3.12. *Let F, S, and $\{\mu_\tau\}$ be as in the statement of Theorem 1.3, except now F may be any number field. Then there exist a finite set of places $P \supset S$ and a finite extension F'/F such that any geometric ρ_λ with good reduction outside S, and with Hodge–Tate cocharacters arising from $\{\mu_\tau\}$ via an embedding $\iota_\lambda : \bar{E} \hookrightarrow \bar{E}_\lambda$, admits a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F, P \cup S_\lambda} \to \widetilde{H}(\bar{E}_\lambda)$ such that the restrictions*

$$\xi(\tilde{\rho}_\lambda) : \Gamma_{F', P \cup S_\lambda} \to \widetilde{Z}^\vee(\bar{E}_\lambda)$$

are equal to the ι_λ-adic realizations of the single (independent of λ and ρ_λ) algebraic Hecke character ψ of $\widetilde{Z}(\mathbb{A}_F)$.

In particular, let $\{\rho_\lambda : \Gamma_{F, S \cup S_\lambda} \to H(\bar{E}_\lambda)\}_\lambda$ be a ramification-compatible system. Then there exist a ramification-compatible system of lifts $\{\tilde{\rho}_\lambda : \Gamma_{F, P \cup S_\lambda} \to \widetilde{H}(\bar{E}_\lambda)\}_\lambda$, and a finite, independent-of-λ extension F'/F such that the restrictions

$$\xi(\tilde{\rho}_\lambda) : \Gamma_{F', P \cup S_\lambda} \to \widetilde{Z}^\vee(\bar{E}_\lambda)$$

form a strongly compatible system.

Proof. We may assume that the number field F is totally imaginary. Consider the lifts $\tilde{\rho}_\lambda : \Gamma_{F, P \cup S_\lambda} \to \widetilde{H}(\bar{E}_\lambda)$ produced by Theorem 1.3. We write $\xi(\tilde{\rho}_\lambda) = \psi_\lambda \cdot \eta_\lambda$, where $\eta_\lambda : \Gamma_{F, P \cup S_\lambda} \to \widetilde{Z}^\vee[M]$ is a finite-order character; the independent-of-λ bound on the order was established within the proof of Theorem 1.3. Moreover, for all $v \in S_\lambda \setminus (S_\lambda \cap P)$, $\tilde{\rho}_\lambda$ and ψ_λ are crystalline at v, so as long as $S_\lambda \cap P$ is empty, η_λ factors through $\Gamma_{F, P} \to \widetilde{Z}^\vee[M]$ (we again use that a finite-order crystalline character is unramified). By the Hermite–Minkowski theorem, there are a finite number of such characters η_λ. For the finite number of bad λ (at which $S_\lambda \cap P \neq \varnothing$), the same finiteness assertion holds. Thus, after a finite base change F'/F, trivializing this finite collection of possible characters η_λ, we see that $\xi(\tilde{\rho}_\lambda)|_{\Gamma_{F', P \cup S_\lambda}} = \psi_\lambda|_{\Gamma_{F', P \cup S_\lambda}}$ for all λ. The second part of the corollary follows since the λ-adic realizations of an abelian L-algebraic representation form a strongly compatible system, as is evident from the construction of ψ_λ, as in, e.g., [Serre 1968]. □

We would like to upgrade this to a compatibility statement not just for the push-forwards $\xi(\tilde{\rho}_\lambda)$, but for the full abelianizations $\tilde{\rho}^{\mathrm{ab}} : \Gamma_{F, P \cup S_\lambda} \to \widetilde{H}^{\mathrm{ab}}(\bar{E}_\lambda)$. Of course, such a result requires first (taking $\widetilde{H} = H$) having the corresponding assertion for the abelianizations $\rho_\lambda^{\mathrm{ab}} : \Gamma_{F, S \cup S_\lambda} \to H^{\mathrm{ab}}(\bar{E}_\lambda)$. Here, however, it is of course false without imposing further conditions on the system $\{\rho_\lambda\}_\lambda$ (see Remark 3.15). There are various conditions we might impose on the ρ_λ to ensure (potential) compatibility of the $\rho_\lambda^{\mathrm{ab}}$. Perhaps most interesting is to restrict the coefficients of $\rho_\lambda^{\mathrm{ab}}$. To that end, we first prove a finiteness result for Galois characters:

Lemma 3.13. *Let F be a number field, and let S be a finite set of places of F. Fix a finite extension E'/E (inside $\overline{E}$), a set $\{m_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ of integers satisfying the Hodge-symmetry condition of Definition 1.1, and an embedding $\iota_\infty : \overline{E} \hookrightarrow \mathbb{C}$. Then there exist a finite extension F'/F, and an algebraic Hecke character α of $\mathbb{A}_{F'}$, such that any geometric character $\omega_\lambda : \Gamma_{F,S\cup S_\lambda} \to \overline{E}_\lambda^\times$*

- *having good reduction outside S;*
- *having labeled Hodge–Tate weights corresponding to $\{m_\tau\}$ via some embedding $\iota_\lambda : \overline{E} \hookrightarrow \overline{E}_\lambda$; and*
- *for which $\omega_\lambda(fr_v)$ belongs to $(E')^\times$ for a density-one set of places v of F;*

will upon restriction $\omega_\lambda|_{\Gamma_{F'}}$ become isomorphic to the ι_λ-adic realization of α.

Proof. We may assume F is totally imaginary. Invoking the Hodge symmetry hypothesis, we apply Lemma 3.1 to produce an algebraic Hecke character α of F whose archimedean components are given in terms of the m_τ, exactly as in Lemma 3.1 (with $\xi(\widetilde{\mu}_\tau) = m_\tau$). Let T denote the finite set of ramified places of α, and let $\mathbb{Q}(\alpha)$ denote the field of coefficients of α (by definition the fixed field of all automorphisms of $\mathbb{C}$ that preserve the nonarchimedean component of α; we will regard $\mathbb{Q}(\alpha)$ as a subfield of $\overline{E}$ via our fixed ι_∞). Thus the ι_λ-adic realizations $\alpha_\lambda : \Gamma_{F,T\cup S_\lambda} \to \overline{E}_\lambda^\times$ have labeled Hodge–Tate weights matching those of ω_λ. Since $\mathbb{Q}(\alpha)$ contains the values $\alpha_\lambda(fr_v)$ for all $v \notin T \cup S_\lambda$, and $\omega_\lambda \alpha_\lambda^{-1} : \Gamma_{F,S\cup T\cup S_\lambda} \to \overline{E}_\lambda^\times$ is finite-order (all of its Hodge–Tate weights are zero), we see that $(\omega_\lambda\alpha_\lambda^{-1})(fr_v)$ belongs to the finite (independent of λ) set $\mu_\infty(E'\mathbb{Q}(\alpha))$ for a density-one set of v. By Čebotarev, the character $\omega_\lambda\alpha_\lambda^{-1}$ takes all of its values in $\mu_\infty(E'\mathbb{Q}(\alpha))$. As long as $S_\lambda \cap (S \cup T)$ is empty, $\omega_\lambda\alpha_\lambda^{-1}$ is moreover unramified at S_λ (because it is crystalline of finite order), so as in Corollary 3.12, there are (again by Hermite–Minkowski) a finite number of such characters $\omega_\lambda\alpha_\lambda^{-1}$. We deduce the existence of a single number field F' over which $\omega_\lambda|_{\Gamma_{F'}} = \alpha_\lambda|_{\Gamma_{F'}}$, for any λ and any ω_λ as in the statement of the lemma. □

We deduce a potential compatibility statement for the full abelianizations $\tilde{\rho}_\lambda^{\mathrm{ab}}$:

Corollary 3.14. *For simplicity, assume that H^{ab} is of multiplicative type. Let F, S, and $\{\mu_\tau\}$ be as in the statement of Theorem 1.3, except with F now allowed to be any number field. Also fix a finite extension E' of E. Then there exist a finite set of places $P \supset S$, a finite extension F'/F, and an algebraic Hecke character β of the split group $\widetilde{D}$ over F' whose dual group over $\overline{E}$ is isomorphic to $(H^{\mathrm{ab}})^0 \otimes_E \overline{E}$ (and we fix such an isomorphism), satisfying the following: if a geometric $\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(\overline{E}_\lambda)$*

- *has good reduction outside* S;
- *has Hodge–Tate cocharacters arising from* $\{\mu_\tau\}$ *via* $\iota_\lambda : \bar{E} \hookrightarrow \bar{E}_\lambda$;
- *and admits, for some faithful representation* r *of* H^{ab}, *a density-one set of places* v *of* F *such that the characteristic polynomial* $\mathrm{ch}(r \circ \tilde{\rho}_\lambda^{\mathrm{ab}})(fr_v)$ *has coefficients in* E';

then there is a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F,P\cup S_\lambda} \to \widetilde{H}(\bar{E}_\lambda)$ *having good reduction outside* P, *such that the restriction* $\tilde{\rho}_\lambda^{\mathrm{ab}} : \Gamma_{F',P\cup S_\lambda} \to \widetilde{H}^{\mathrm{ab}}(\bar{E}_\lambda)$ *is equal to the* ι_λ*-adic realization* β_λ *of* β.

In particular, let $\{\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(\bar{E}_\lambda)\}_\lambda$ *be a ramification-compatible system, and assume that for some faithful representation* r *of* H^{ab}, *some number field* E', *and for almost all* λ, *there is a density-one set of places* v *of* F *such that the characteristic polynomial* $\mathrm{ch}(r \circ \tilde{\rho}_\lambda^{\mathrm{ab}})(fr_v)$ *has coefficients in* E'. *Then there is a ramification-compatible system* $\tilde{\rho}_\lambda : \Gamma_{F,P\cup S_\lambda} \to \widetilde{H}(\bar{E}_\lambda)$ *lifting* ρ_λ, *and a finite extension* F'/F *such that*

$$\tilde{\rho}_\lambda^{\mathrm{ab}}|_{\Gamma_{F'}} : \Gamma_{F',P\cup S_\lambda} \to \widetilde{H}^{\mathrm{ab}}(\bar{E}_\lambda)$$

forms a strongly compatible system.

Proof. The proof follows familiar lines. Since C is central, the abelianization $\widetilde{H}^{\mathrm{ab}}$ is simply $\widetilde{H}/H_1^{\mathrm{der}}$, so there is a natural map

$$f : \widetilde{H}^{\mathrm{ab}} \to \widetilde{H}/H_1 \times H/\mathrm{im}(H_1^{\mathrm{der}}) = \widetilde{Z}^\vee \times H^{\mathrm{ab}}$$

under which $\tilde{\rho}_\lambda^{\mathrm{ab}}$ pushes forward to $(\xi(\tilde{\rho}_\lambda), \rho_\lambda^{\mathrm{ab}})$. (We have chosen $\{\tilde{\rho}_\lambda\}_\lambda$ as in Theorem 1.3 and Corollary 3.12, of course.) First we claim that a conclusion analogous to that of the corollary holds for the pair $(\xi(\tilde{\rho}_\lambda), \rho_\lambda^{\mathrm{ab}})$, and certainly it suffices to check this independently for the two components. The assertion for $\xi(\tilde{\rho}_\lambda)$ is Corollary 3.12, and for $\rho_\lambda^{\mathrm{ab}}$ it follows easily from Lemma 3.13 (first reduce, by a finite base change, to the case where H^{ab} is connected, using the fact that $\pi_0(H^{\mathrm{ab}})$ is of course finite and independent of λ). Thus, letting D denote a split torus whose dual group is identified with $(H^{\mathrm{ab}})^0$, there exists a finite extension F_1/F such that $f(\tilde{\rho}_\lambda^{\mathrm{ab}})|_{\Gamma_{F_1,P\cup S_\lambda}}$ is the ι_λ-adic realization of a Hecke character (which does not depend on λ or ρ_λ) of $\widetilde{Z} \times D$.

Now suppose that β is a Hecke character of $\widetilde{D}$ for which the ι_λ-adic realization $\beta_\lambda : \Gamma_{F,T\cup S_\lambda} \to (\widetilde{H}^{\mathrm{ab}})^0(\bar{E}_\lambda)$ has labeled Hodge–Tate cocharacters matching those of $\tilde{\rho}_\lambda^{\mathrm{ab}}$. Since

$$f(\tilde{\rho}_\lambda^{\mathrm{ab}} \cdot \beta_\lambda^{-1})|_{\Gamma_{F_1,P\cup T\cup S_\lambda}}$$

is automorphic (independently of λ, ρ_λ) of finite order, it is trivial after a finite base change F_2/F_1. Now observe that the kernel of f is finite, so $(\tilde{\rho}_\lambda^{\mathrm{ab}} \cdot \beta_\lambda^{-1})|_{\Gamma_{F_2,P\cup T\cup S_\lambda}}$ has finite order, bounded only in terms of $\#\ker(f)$, and is crystalline away from $P \cup T$;

as before, we find a further finite extension F_3/F_2 such that $(\tilde{\rho}_\lambda^{\mathrm{ab}} \cdot \beta_\lambda^{-1})|_{\Gamma_{F_3}} = 1$. The conclusion of the corollary then holds with $F' = F_3$. □

Remark 3.15.
- It does not suffice to ask for a fixed number field E such that all $\rho_\lambda^{\mathrm{ab}}$ are valued in $H^{\mathrm{ab}}(E_\lambda)$. For instance, taking $F = \mathbb{Q}$ and $S = \{p\}$, and for all n choosing a prime $\ell_n \equiv 1 \pmod{\varphi(p^n)}$, we can define $\rho_{\ell_n} : \Gamma_{\mathbb{Q},\{p\}} \to \mathbb{Q}_{\ell_n}^\times$ as the composition of the mod p^n cyclotomic character with an inclusion $(\mathbb{Z}/p^n\mathbb{Z})^\times \hookrightarrow \mu_{\ell_n-1} \hookrightarrow \mathbb{Q}_{\ell_n}^\times$, and for all $\ell \notin \{\ell_n\}_n$ we can take ρ_ℓ to be the trivial character. Then $\{\rho_\ell : \Gamma_{\mathbb{Q},\{p\}} \to \mathbb{Q}_\ell^\times\}_\ell$ is an abelian, ramification-compatible system that does not become a strongly compatible system after any finite base change.
- Having only hypothesized ramification-compatibility for the $\{\rho_\lambda\}_\lambda$, we cannot hope for the stronger conclusion that the $\{\tilde{\rho}_\lambda^{\mathrm{ab}}\}_\lambda$ form a strongly compatible system over F itself.

General multiplicative-type quotients. In fact, the argument of Theorem 1.3 directly implies the main theorem of [Wintenberger 1995], as well as a generalization to lifting through quotients where the kernel is central of multiplicative type. We thus obtain an essentially different proof (and generalization) of Wintenberger's result. In this section, we briefly describe how this works.

Corollary 3.16 (Wintenberger). *Let $H_1 \to H$ be a central isogeny of linear algebraic groups over E, and let S be a finite set of places of F. Then there exist a finite extension F'/F and a finite set of places $P \supset S$ of F such that any geometric representation*

$$\rho_\lambda : \Gamma_{F,S\cup S_\lambda} \to H(E_\lambda)$$

having

- *good reduction outside S, and*
- *labeled Hodge–Tate cocharacters that lift to H_1,*

lifts to a geometric representation $\rho'_\lambda : \Gamma_{F',P\cup S_\lambda} \to H_1(E_\lambda)$, which moreover has good reduction outside P.

Proof. We begin by replacing F by a finite extension F_0 such that image of $\rho_\lambda|_{\Gamma_{F_0}}$ is contained in the image of $H_1(E_\lambda) \to H(E_\lambda)$. That such an extension, depending only on $H_1 \to H$, S, and F, exists follows as in [Wintenberger 1995, 2.3.2], and we do not repeat the argument. We note, though, that making this construction in an independent-of-λ fashion already uses liftability of the Hodge–Tate cocharacters. (If we were not concerned with preserving E_λ-rationality of the lift, then we could skip this step.) It is then possible to build an obstruction class $\mathcal{O}(\rho_\lambda) \in H^2(\Gamma_{F_0,S\cup S_\lambda}, \ker(H_1 \to H)(E_\lambda))$ via a topological lift ρ'_λ to $H_1(E_\lambda)$.

Embed $\ker(H_1 \to H) \otimes_E \overline{E}$ into a torus C, and with the kernel embedded antidiagonally, form the new group $\widetilde{H} = (H_1 \times C)/\ker(H_1 \to H)$. The surjection $\widetilde{H} \to H$ now has kernel equal to a central torus C, and as before we let $\widetilde{Z}^\vee$ be the (torus) quotient $\widetilde{H}/H_1$. By hypothesis, we can lift the Hodge–Tate cocharacters of ρ_λ to H_1; when pushed forward to $\widetilde{Z}^\vee$, these lifts are of course trivial. Thus, in the notation of Lemma 3.1, we may take the trivial Hecke character $\psi = 1$ of $\widetilde{Z}(\mathbb{A}_{F_0})$. For topological lifts ψ'_λ to $C(E_\lambda)$ (as in Lemma 3.3) of the (trivial) λ-adic realizations ψ_λ, we may of course also take $\psi'_\lambda = 1$. Theorem 1.3 then produces a finite set of primes $P \supset S$ and an integer M, both only depending on $H_1 \to H$, S, and F, and a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F_0, P\cup S_\lambda} \to H_1(E_\lambda) \cdot C[M]$ such that $\tilde{\rho}_\lambda$ has good reduction outside P. (The assertion that $\tilde{\rho}_\lambda$ is valued in the subset $H_1(E_\lambda) \cdot C[M]$ of $\widetilde{H}(\overline{E}_\lambda)$ follows from the explicit description of $\tilde{\rho}_\lambda$, since ρ'_λ lands in $H_1(E_\lambda)$, and ψ'_λ is trivial.) For all λ for which $S_\lambda \cap P = \varnothing$, $\xi(\tilde{\rho}_\lambda) : \Gamma_{F_0, P\cup S_\lambda} \to \widetilde{Z}^\vee[M]$ is also unramified at S_λ, and all such characters are trivialized by a common finite extension F_1/F_0. For the finite number of λ such that $S_\lambda \cap P$ is nonempty, we can again trivialize the possible $\xi(\tilde{\rho}_\lambda)$ by restricting to a common finite extension F_2/F_0. Taking $F' = F_1 F_2$, all $\tilde{\rho}_\lambda|_{\Gamma_{F', P\cup S_\lambda}}$ land in $H_1(E_\lambda)$, proving the corollary. □

Remark 3.17. For Wintenberger's result, take $E = \mathbb{Q}$. He also shows [1995, 2.3.6] that there is a second finite extension F''/F' (only depending on $H_1 \to H$, F, and S) such that any two lifts ρ'_λ as in the corollary become equal after restriction to $\Gamma_{F''}$. This refinement similarly follows in our setup, but there is no need to repeat Wintenberger's argument.

Here is the more general version with multiplicative-type kernels. Note that, as with Theorem 1.3, but unlike Corollary 3.16, it makes use of a "Hodge symmetry" hypothesis.

Corollary 3.18. *Let $H' \to H$ be a surjection of linear algebraic groups over E whose kernel is central and of multiplicative type. Let F be a number field, and let S be a finite set of places of F containing the archimedean places. Fix a set of cocharacters $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ as in part (2) of Definition 1.1, and moreover, assume that each μ_τ lifts to a cocharacter of H'.*

Then there exist a finite set of places $P \supset S$, and a finite extension F'/F, such that any geometric representation $\rho_\lambda : \Gamma_{F, S\cup S_\lambda} \to H(\overline{E}_\lambda)$ having good reduction outside S, and whose Hodge–Tate cocharacters arise from the set $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$ via some embedding $\overline{E} \hookrightarrow \overline{E}_\lambda$, admits a geometric lift $\tilde{\rho}_\lambda : \Gamma_{F', P\cup S_\lambda} \to H'(\overline{E}_\lambda)$ having good reduction outside P.

In particular, if $\{\rho_\lambda : \Gamma_{F, S\cup S_\lambda} \to H(\overline{E}_\lambda)\}_\lambda$ is a ramification-compatible system with Hodge cocharacter $\{\mu_\tau\}_{\tau : F \hookrightarrow \overline{E}}$, then there exist a finite set of places $P \supset S$, a finite extension F'/F, and lifts $\tilde{\rho}_\lambda : \Gamma_{F', P\cup S_\lambda} \to H'(\overline{E}_\lambda)$ such that $\{\tilde{\rho}_\lambda\}_\lambda$ is a ramification-compatible system.

Proof. As in the proofs of Theorem 1.3 and Corollary 3.16, we construct an isogeny complement $H_1 \subset H$ to $\ker(H' \to H)$, as well as an enlargement $\widetilde{H} \supset H'$ surjecting onto H with a central torus kernel. We then run the argument of Theorem 1.3, starting from lifts $\{\mu'_\tau\}$ to H' of the Hodge cocharacters: the Hecke character ψ (in the notation of that proof) then constructed has λ-adic realizations that push-forward to finite-order characters $\psi_\lambda : \Gamma_{F,T \cup S_\lambda} \to \widetilde{H}/H'(\overline{E}_\lambda)$, and from here it is easy to proceed; we omit the details, since the argument will by now be familiar. $\square$

References

[André 1996a] Y. André, "On the Shafarevich and Tate conjectures for hyper-Kähler varieties", *Math. Ann.* **305**:2 (1996), 205–248. MR Zbl

[André 1996b] Y. André, "Pour une théorie inconditionnelle des motifs", *Inst. Hautes Études Sci. Publ. Math.* **83** (1996), 5–49. MR Zbl

[Artin and Tate 1968] E. Artin and J. Tate, *Class field theory*, W. A. Benjamin, New York, 1968. MR Zbl

[Barnet-Lamb et al. 2014] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Potential automorphy and change of weight", *Ann. of Math.* (2) **179**:2 (2014), 501–609. MR Zbl

[Chai et al. 2014] C.-L. Chai, B. Conrad, and F. Oort, *Complex multiplication and lifting problems*, Mathematical Surveys and Monographs **195**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Chevalley 1951] C. Chevalley, "Deux théorèmes d'arithmétique", *J. Math. Soc. Japan* **3** (1951), 36–44. MR Zbl

[Conrad 2011] B. Conrad, "Lifting global representations with local properties", preprint, 2011, available at http://math.stanford.edu/~conrad/papers/locchar.pdf.

[Deligne 1972] P. Deligne, "La conjecture de Weil pour les surfaces $K3$", *Invent. Math.* **15** (1972), 206–226. MR Zbl

[Deligne 1977] P. Deligne, *Cohomologie étale*, Lecture Notes in Math. **569**, Springer, 1977. MR Zbl

[Deligne and Milne 1982] P. Deligne and J. S. Milne, "Tannakian categories", pp. 101–228 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. **900**, Springer, 1982. MR Zbl

[Faltings 1989] G. Faltings, "Crystalline cohomology and p-adic Galois-representations", pp. 25–80 in *Algebraic analysis, geometry, and number theory* (Baltimore, 1988), edited by J.-I. Igusa, Johns Hopkins University Press, 1989. MR Zbl

[Kret and Shin 2016] A. Kret and S. W. Shin, "Galois representations for general symplectic groups", preprint, 2016. arXiv

[Kuga and Satake 1967] M. Kuga and I. Satake, "Abelian varieties attached to polarized K_3-surfaces", *Math. Ann.* **169** (1967), 239–242. MR Zbl

[Neukirch et al. 2000] J. Neukirch, A. Schmidt, and K. Wingberg, *Cohomology of number fields*, Grundlehren der Math. Wissenschaften **323**, Springer, 2000. MR Zbl

[Patrikis 2015] S. Patrikis, "On the sign of regular algebraic polarizable automorphic representations", *Math. Ann.* **362**:1-2 (2015), 147–171. MR Zbl

[Patrikis 2016a] S. Patrikis, "Generalized Kuga–Satake theory and rigid local systems, I: The middle convolution", pp. 368–392 in *Recent advances in Hodge theory*, edited by M. Kerr and G. Pearlstein, London Math. Society Lecture Note Series **427**, Cambridge University Press, 2016. MR Zbl

[Patrikis 2016b] S. Patrikis, "Generalized Kuga–Satake theory and rigid local systems, II: Rigid Hecke eigensheaves", *Algebra Number Theory* **10**:7 (2016), 1477–1526. MR Zbl

[Patrikis 2016c] S. Patrikis, "Variations on a theorem of Tate", 2016, available at http://tinyurl.com/tatevar. To appear in *Mem. Amer. Math. Soc.*

[Serre 1968] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, W. A. Benjamin, New York, 1968. MR Zbl

[Serre 1977] J.-P. Serre, "Modular forms of weight one and Galois representations", pp. 193–268 in *Algebraic number fields: L-functions and Galois properties* (Durham, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR Zbl

[Weil 1956] A. Weil, "On a certain type of characters of the idèle-class group of an algebraic number-field", pp. 1–7 in *Proceedings of the international symposium on algebraic number theory* (Tokyo/Nikko, 1955), Sci. Council Japan, Tokyo, 1956. MR Zbl

[Wintenberger 1995] J.-P. Wintenberger, "Relèvement selon une isogénie de systèmes l-adiques de représentations galoisiennes associés aux motifs", *Invent. Math.* **120**:2 (1995), 215–240. MR Zbl

Communicated by Christopher Skinner
Received 2017-04-30 Revised 2017-08-08 Accepted 2017-09-06

patrikis@math.utah.edu *Department of Mathematics, University of Utah, Salt Lake City, UT, United States*

ALGEBRA AND NUMBER THEORY 11:10 (2017)
dx.doi.org/10.2140/ant.2017.11.2425

Remarks on the arithmetic fundamental lemma

Chao Li and Yihang Zhu

W. Zhang's arithmetic fundamental lemma (AFL) is a conjectural identity between the derivative of an orbital integral on a symmetric space with an arithmetic intersection number on a unitary Rapoport–Zink space. In the minuscule case, Rapoport, Terstiege and Zhang have verified the AFL conjecture via explicit evaluation of both sides of the identity. We present a simpler way for evaluating the arithmetic intersection number, thereby providing a new proof of the AFL conjecture in the minuscule case.

1. Introduction

1.1. ***Zhang's arithmetic fundamental lemma.*** The *arithmetic Gan–Gross–Prasad conjectures* (arithmetic GGP) generalize the celebrated Gross–Zagier formula to higher-dimensional Shimura varieties [Gan et al. 2012, §27; Zhang 2012, §3.2]. The arithmetic fundamental lemma (AFL) conjecture arises from Zhang's relative trace formula approach for establishing the arithmetic GGP for the group $U(1, n-2) \times U(1, n-1)$. It relates a derivative of orbital integrals on symmetric spaces to an arithmetic intersection number of cycles on unitary Rapoport–Zink spaces,

$$O'(\gamma, \mathbf{1}_{S_n(\mathbb{Z}_p)}) = -\omega(\gamma)\langle \Delta(\mathcal{N}_{n-1}), (\mathrm{id} \times g)\Delta(\mathcal{N}_{n-1})\rangle. \tag{1.1.0.1}$$

For the precise definitions of quantities appearing in the identity, see [Rapoport et al. 2013, Conjecture 1.2]. The left-hand side of (1.1.0.1) is known as the *analytic side* and the right-hand side is known as the *arithmetic-geometric* side. The AFL conjecture has been verified for $n = 2, 3$ [Zhang 2012], and for general n in the minuscule case (in the sense that g satisfies a certain minuscule condition) by Rapoport, Terstiege and Zhang [2013]. In all these cases, the identity (1.1.0.1) is proved via explicit evaluation of both sides. When g satisfies a certain inductive condition, Mihatsch [2016] has recently developed a recursive algorithm which

MSC2010: primary 11G18; secondary 14G17, 22E55.
Keywords: arithmetic Gan–Gross–Prasad conjectures, arithmetic fundamental lemmas, Rapoport–Zink spaces, special cycles.

reduces the identity (1.1.0.1) to smaller n, thus establishing some new cases of the AFL conjecture.

In the minuscule case, the evaluation of the analytic side is relatively straightforward. The bulk of [Rapoport et al. 2013] is devoted to a highly nontrivial evaluation of the arithmetic-geometric side, which is truly a tour de force. Our main goal in this short note is to present a new (and arguably simpler) way to evaluate the arithmetic-geometric side in [Rapoport et al. 2013], henceforth abbreviated [RTZ].

1.2. ***The main results.*** Let p be an odd prime. Let $F=\mathbb{Q}_p$, $k=\overline{\mathbb{F}}_p$, $W=W(k)$ and $K=W[1/p]$. Let σ be the p-Frobenius acting on $\overline{\mathbb{F}}_p$, W and K. Let $E=\mathbb{Q}_{p^2}$ be the unramified quadratic extension of F. The *unitary Rapoport–Zink space* $\mathcal{N}_n$ is the formal scheme over $\operatorname{Spf} W$ parametrizing deformations up to quasi-isogeny of height 0 of unitary p-divisible groups of signature $(1, n-1)$ (definitions recalled in Section 2.1). Fix $n\geq 2$ and write $\mathcal{N}=\mathcal{N}_n$ and $\mathcal{M}=\mathcal{N}_{n-1}$ for short. There is a natural closed immersion $\delta:\mathcal{M}\to\mathcal{N}$. Denote by $\Delta\subseteq\mathcal{M}\times_W\mathcal{N}$ the image of $(\mathrm{id},\delta):\mathcal{M}\to\mathcal{M}\times_W\mathcal{N}$, known as the (local) *diagonal cycle* or *GGP cycle* on $\mathcal{M}\times_W\mathcal{N}$.

Let C_{n-1} be a nonsplit σ-Hermitian E-space of dimension $n-1$. Let $C_n=C_{n-1}\oplus Eu$ (where the direct sum is orthogonal and u has norm 1) be a nonsplit σ-Hermitian E-space of dimension n. The unitary group $J=\mathrm{U}(C_n)$ acts on $\mathcal{N}$ in a natural way (see Section 2.2). Let $g\in J(\mathbb{Q}_p)$. The arithmetic-geometric side of the AFL conjecture (1.1.0.1) concerns the arithmetic intersection number of the diagonal cycle Δ and its translate by $\mathrm{id}\times g$, defined as

$$\langle\Delta,(\mathrm{id}\times g)\Delta\rangle:=\log p\cdot\chi(\mathcal{M}\times_W\mathcal{N},\mathcal{O}_\Delta\otimes^{\mathbb{L}}\mathcal{O}_{(\mathrm{id}\times g)\Delta}).$$

When Δ and $(\mathrm{id}\times g)\Delta$ intersect properly, namely when the formal scheme

$$\Delta\cap(\mathrm{id}\times g)\Delta\cong\delta(\mathcal{M})\cap\mathcal{N}^g \tag{1.2.0.1}$$

is an Artinian scheme (where $\mathcal{N}^g$ denotes the fixed points of g), the intersection number is simply $\log p$ times the W-length of the Artinian scheme (1.2.0.1).

Recall that $g\in J(\mathbb{Q}_p)$ is called *regular semisimple* if

$$L(g):=\mathcal{O}_E\cdot u+\mathcal{O}_E\cdot gu+\cdots+\mathcal{O}_E\cdot g^{n-1}u$$

is an $\mathcal{O}_E$-lattice in C_n. In this case, the *invariant* of g is the unique sequence of integers

$$\mathrm{inv}(g):=(r_1\geq r_2\geq\cdots\geq r_n)$$

characterized by the condition that there exists a basis $\{e_i\}$ of the lattice $L(g)$ such that $\{p^{-r_i}e_i\}$ is a basis of the dual lattice $L(g)^\vee$. It turns out that the "bigger" $\mathrm{inv}(g)$ is, the more difficult it is to compute the intersection. With this in mind, recall that a regular semisimple element g is called *minuscule* if $r_1=1$ and $r_n\geq 0$ (equivalently, $pL(g)^\vee\subseteq L(g)\subseteq L(g)^\vee$). In this minuscule case, the intersection turns out to be proper, and one of the main results of [RTZ] is an explicit formula for the W-length of (1.2.0.1) at each of its k-point.

To state the formula, assume g is regular semisimple and minuscule, and suppose $\mathcal{N}^g$ is nonempty. Then g stabilizes both $L(g)^\vee$ and $L(g)$ and thus acts on the $\mathbb{F}_{p^2}$-vector space $L(g)^\vee/L(g)$. Let $P(T) \in \mathbb{F}_{p^2}[T]$ be the characteristic polynomial of g acting on $L(g)^\vee/L(g)$. For any irreducible polynomial $R(T) \in \mathbb{F}_{p^2}[T]$, we denote its multiplicity in $P(T)$ by $m(R(T))$ and define its *reciprocal* by

$$R^*(T) := T^{\deg R(T)} \cdot \sigma\big(R\big(\tfrac{1}{T}\big)\big).$$

We say $R(T)$ is *self-reciprocal* if $R(T) = R^*(T)$. By [RTZ, 8.1], if $(\delta(M) \cap \mathcal{N}^g)(k)$ is nonempty, then $P(T)$ has a unique self-reciprocal monic irreducible factor $Q(T) \mid P(T)$ such that $m(Q(T))$ is odd. We denote

$$c := \frac{1}{2}(m(Q(T)) + 1).$$

Then $1 \leq c \leq \frac{1}{2}(n+1)$. Now we are ready to state the intersection length formula.

Theorem A [RTZ, Theorem 9.5]. *Assume g is regular semisimple and minuscule. Assume $p > c$. Then for any $x \in (\delta(M) \cap \mathcal{N}^g)(k)$, the complete local ring of $\delta(M) \cap \mathcal{N}^g$ at x is isomorphic to $k[X]/X^c$, and hence has W-length equal to c.*

We will present a simpler proof of Theorem A in Theorem 4.3.5. Along the way, we will also give a simpler proof of the following Theorem B in Corollary 3.2.3, which concerns minuscule special cycles (recalled in Section 2.10) on unitary Rapoport–Zink spaces and may be of independent interest.

Theorem B [RTZ, Theorems 9.4 and 10.1]. *Let $\boldsymbol{v} = (v_1, \dots, v_n)$ be an n-tuple of vectors in C_n. Assume it is minuscule in the sense that $L(\boldsymbol{v}) := \operatorname{span}_{\mathcal{O}_E} \boldsymbol{v}$ is an $\mathcal{O}_E$-lattice in C_n satisfying $pL(\boldsymbol{v})^\vee \subseteq L(\boldsymbol{v}) \subseteq L(\boldsymbol{v})^\vee$. Let $\mathcal{Z}(\boldsymbol{v}) \subseteq \mathcal{N}$ be the associated special cycle. Then $\mathcal{Z}(\boldsymbol{v})$ is a reduced k-scheme.*

1.3. ***Novelty of the proof.*** The original proofs of Theorems A and B form the technical heart of [RTZ] and occupy its two sections §10–§11. As explained below, our new proofs presented here have the merit of being much shorter and more conceptual.

1.3.1. *Theorem A.* The original proof of Theorem A uses Zink's theory of windows to compute the local equations of (1.2.0.1). It requires explicitly writing down the window of the universal deformation of p-divisible groups and solving quite involved linear algebra problems. Theorem B ensures that the intersection is entirely concentrated in the special fiber so that each local ring has the form $k[X]/X^\ell$. The assumption $p > c$ ensures $\ell < p$ so that the ideal of local equations is *admissible* (see the last paragraph of [RTZ, p. 1661]), which is crucial in order to construct the frames for the relevant windows needed in Zink's theory.

Our new proof of Theorem A does not use Zink's theory and involves little explicit computation. Our key observation is that Theorem B indeed allows us to identify the intersection (1.2.0.1) as the fixed point scheme $\mathcal{V}(\Lambda)^{\bar{g}}$ of a finite order automorphism $\bar{g}$ on a generalized Deligne–Lusztig variety $\mathcal{V}(\Lambda)$ (Section 4.1), which becomes purely an algebraic geometry problem over the *residue field* k. When $p > c$, it further simplifies to a more elementary problem of determining the fixed point scheme of a finite order automorphism $\bar{g} \in \mathrm{GL}_{d+1}(k)$ on a projective space $\mathbb{P}^d$ over k (Section 4.2). This elementary problem has an answer in terms of the sizes of the Jordan blocks of $\bar{g}$ (Lemma 4.3.4), which explains conceptually why the intersection multiplicity should be equal to c. Notice that our method completely avoids computation within Zink's theory, and it would be interesting to explore the possibility of removing the assumption $p > c$ using this method.

1.3.2. *Theorem B.* The original proof of Theorem B relies on showing two things (by [RTZ, Lemma 10.2]): (1) the minuscule special cycle $\mathcal{Z}(\boldsymbol{v})$ has no W/p^2-points and (2) its special fiber $\mathcal{Z}(\boldsymbol{v})_k$ is regular. Step (1) is relatively easy using Grothendieck–Messing theory. Step (2) is more difficult: for super-general points x on $\mathcal{Z}(\boldsymbol{v})_k$, the regularity is shown by explicitly computing the local equation of $\mathcal{Z}(\boldsymbol{v})_k$ at x using Zink's theory; for points which are not super-general, the regularity is shown using induction and reduces to the regularity of certain special divisors, whose local equations can again be explicitly computed using Zink's theory.

Our new proof of Theorem B does not use Zink's theory either and involves little explicit computation. Our key observation is that to show both (1) and (2) it suffices to consider the thickenings $\mathcal{O}$ of k which are objects of the crystalline site of k. These $\mathcal{O}$-points of $\mathcal{Z}(\boldsymbol{v})$ then can be understood using only Grothendieck–Messing theory (Theorem 3.1.3). We prove a slight generalization of (1) which applies to possibly nonminuscule special cycles (Corollary 3.2.1). We then prove the tangent space of the minuscule special cycle $\mathcal{Z}(\boldsymbol{v})_k$ has the expected dimension (Corollary 3.2.2). The desired regularity (2) follows immediately.

1.3.3. Our new proofs are largely inspired by our previous work on arithmetic intersections on GSpin Rapoport–Zink spaces [Li and Zhu 2017]. The GSpin Rapoport–Zink spaces considered there are not of PEL type, which makes them technically more complicated. So the unitary case treated here can serve as a guide to [Li and Zhu 2017]. We have tried to indicate similarities between certain statements and proofs, for both clarity and the convenience of the readers.

1.4. ***Structure of the paper.*** In Section 2, we recall necessary backgrounds on unitary Rapoport–Zink spaces and the formulation of the arithmetic intersection problem. In Section 3, we study the local structure of the minuscule special cycles and prove Theorem B. In Section 4, we provide an alternative moduli interpretation of the generalized Deligne–Lusztig variety $\mathcal{V}(\Lambda)$ and prove Theorem A.

2. Unitary Rapoport–Zink spaces

In this section we review the structure of unitary Rapoport–Zink spaces. We refer to [Vollaard 2010; Vollaard and Wedhorn 2011; Kudla and Rapoport 2011] for the proofs of these facts.

2.1. ***Unitary Rapoport–Zink spaces.*** Let p be an odd prime. Let $F = \mathbb{Q}_p$, $k = \overline{\mathbb{F}}_p$, $W = W(k)$ and $K = W[1/p]$. Let σ be the p-Frobenius acting on $\overline{\mathbb{F}}_p$, and we also denote by σ the canonical lift of the p-Frobenius to W and K. For any $\mathbb{F}_p$-algebra R, we also denote by σ the Frobenius $x \mapsto x^p$ on R.

Let $E = \mathbb{Q}_{p^2}$ be the unramified quadratic extension of F. Fix a $\mathbb{Q}_p$-algebra embedding $\phi_0 : \mathcal{O}_E \hookrightarrow W$ and denote by ϕ_1 the embedding $\sigma \circ \phi_0 : \mathcal{O}_E \hookrightarrow W$. The embedding ϕ_0 induces an embedding between the residue fields $\mathbb{F}_{p^2} \hookrightarrow k$, which we shall think of as the natural embedding. For any $\mathcal{O}_E$-module Λ we shall write Λ_W for $\Lambda \otimes_{\mathcal{O}_E, \phi_0} W$.

Let r and s be positive integers and let $n = r + s$. We denote by $\mathcal{N}_{r,s}$ the unitary Rapoport–Zink spaces of signature (r, s), a formally smooth formal W-scheme, parametrizing deformations up to quasi-isogeny of height 0 of unitary p-divisible groups of signature (r, s). More precisely, for a W-scheme S, a *unitary p-divisible groups* of signature (r, s) over S is a triple (X, ι, λ), where

(1) X is a p-divisible group of dimension n and height $2n$ over S,

(2) $\iota : \mathcal{O}_E \to \mathrm{End}(X)$ is an action satisfying the signature (r, s) condition, i.e., for $\alpha \in \mathcal{O}_E$,
$$\mathrm{char}\big(\iota(\alpha) : \mathrm{Lie}\, X\big)(T) = (T - \phi_0(\alpha))^r (T - \phi_1(\sigma))^s \in \mathcal{O}_S[T],$$

(3) $\lambda : X \to X^\vee$ is a principal polarization such that the associated Rosati involution induces $\alpha \mapsto \sigma(\alpha)$ on $\mathcal{O}_E$ via ι.

Over k, there is a unique such triple (X, ι, λ) such that X is supersingular, up to $\mathcal{O}_E$-linear isogeny preserving the polarization up to scalars. Fix such a framing triple and denote it by $(\mathbb{X}, \iota_{\mathbb{X}}, \lambda_{\mathbb{X}})$.

Let Nilp_W be the category of W-schemes on which p is locally nilpotent. Then the *unitary Rapoport–Zink space* $\mathcal{N}_{r,s}$ represents the functor $\mathrm{Nilp}_W \to \mathbf{Sets}$ which sends $S \in \mathrm{Nilp}_W$ to the set of isomorphism classes of quadruples $(X, \iota, \lambda, \rho)$, where (X, ι, λ) is a unitary p-divisible group over S of signature (r, s) and $\rho : X \times_S S_k \to \mathbb{X} \times_k S_k$ is an $\mathcal{O}_E$-linear quasi-isogeny of height zero which respects λ and $\lambda_{\mathbb{X}}$ up to a scalar $c(\rho) \in \mathcal{O}_F^\times = \mathbb{Z}_p^\times$ (i.e., $\rho^\vee \circ \lambda_{\mathbb{X}} \circ \rho = c(\rho) \cdot \lambda$).

In the following we denote $\mathcal{N} := \mathcal{N}_{1,n-1}$, $\mathcal{M} := \mathcal{N}_{1,n-2}$ and $\overline{\mathcal{N}}_0 := \mathcal{N}_{0,1} \cong \mathrm{Spf}\, W$. They have relative dimension $n-1$, n and 0 over $\mathrm{Spf}\, W$ respectively. We denote by $\overline{\mathbb{Y}} = (\overline{\mathbb{Y}}, \iota_{\overline{\mathbb{Y}}}, \lambda_{\overline{\mathbb{Y}}})$ the framing object for $\overline{\mathcal{N}}_0$ and denote by $\overline{Y} = (\overline{Y}, \iota_{\overline{Y}}, \lambda_{\overline{Y}})$ the universal p-divisible group over $\overline{\mathcal{N}}_0$. We may and shall choose framing objects $\mathbb{X} = (\mathbb{X}, \iota_{\mathbb{X}}, \lambda_{\mathbb{X}})$ and $\mathbb{X}^\flat = (\mathbb{X}^\flat, \iota_{\mathbb{X}^\flat}, \lambda_{\mathbb{X}^\flat})$ for $\mathcal{N}$ and $\mathcal{M}$ respectively such that

$$\mathbb{X} = \mathbb{X}^\flat \times \overline{\mathbb{Y}}$$

as unitary p-divisible groups.

2.2. ***The group J.*** The covariant Dieudonné module $M = \mathbb{D}(\mathbb{X})$ of the framing unitary p-divisible group is a free W-module of rank $2n$ together with an $\mathcal{O}_E$-action (induced by ι) and a perfect symplectic W-bilinear form $\langle\cdot,\cdot\rangle : M \times M \to W$ (induced by λ), see [Vollaard and Wedhorn 2011, §2.3]. Let $N = M \otimes_W K$ be the associated isocrystal and extend $\langle\cdot,\cdot\rangle$ to N bilinearly. Let F and V be the usual operators on N. We have

$$\langle Fx, Fy\rangle = p\sigma(\langle x, y\rangle), \quad \forall x, y \in N. \tag{2.2.0.1}$$

The E-action decomposes N into a direct sum of two K-vector spaces of dimension n,

$$N = N_0 \oplus N_1, \tag{2.2.0.2}$$

where the action of E on N_i is induced by the embedding ϕ_i. Both N_0 and N_1 are totally isotropic under the symplectic form. The operator F is of degree one and induces a σ-linear bijection $N_0 \xrightarrow{\sim} N_1$. Since the isocrystal N is supersingular, the degree 0 and σ^2-linear operator

$$\Phi = V^{-1}F = p^{-1}F^2$$

has all slopes zero [Kudla and Rapoport 2011, §2.1]. We have a K-vector space N_0 together with a σ^2-linear automorphism Φ.[1] The space of fixed points

$$C = N_0^{\Phi}$$

is an E-vector space of dimension n and $N_0 = C \otimes_{E,\phi_0} K$. Fix $\delta \in \mathcal{O}_E^\times$ such that $\sigma(\delta) = -\delta$. Define a nondegenerate σ-sesquilinear form on N_0 by

$$\{x, y\} := (p\delta)^{-1}\langle x, Fy\rangle. \tag{2.2.0.3}$$

Using (2.2.0.1) it is easy to see that

$$\sigma(\{x, y\}) = \{\Phi y, x\}, \quad \forall x, y \in N_0. \tag{2.2.0.4}$$

In particular, when restricted to C, the form $\{\cdot,\cdot\}$ is σ-Hermitian, namely

$$\sigma(\{x, y\}) = \{y, x\}, \quad \forall x, y \in C. \tag{2.2.0.5}$$

[1] Such a pair (N_0, Φ) is sometimes called a relative isocrystal.

In fact, $(C, \{\cdot,\cdot\})$ is the unique (up to isomorphism) nondegenerate nonsplit σ-Hermitian E-space of dimension n. Let $J = \mathrm{U}(C)$ be the unitary group of $(C, \{\cdot,\cdot\})$. It is an algebraic group over $F = \mathbb{Q}_p$. By Dieudonné theory, the group $J(\mathbb{Q}_p)$ can be identified with the automorphism group of the framing unitary p-divisible group $(\mathbb{X}, \iota_{\mathbb{X}}, \lambda_{\mathbb{X}})$ and hence acts on the Rapoport–Zink space $\mathcal{N}$.

2.3. ***Special homomorphisms.*** By definition, the *space of special homomorphisms* is the $\mathcal{O}_E$-module $\mathrm{Hom}_{\mathcal{O}_E}(\overline{\mathbb{Y}}, \mathbb{X})$. There is a natural $\mathcal{O}_E$-valued σ-Hermitian form on $\mathrm{Hom}_{\mathcal{O}_E}(\overline{\mathbb{Y}}, \mathbb{X})$ given by

$$(x, y) \mapsto \lambda_{\overline{\mathbb{Y}}}^{-1} \circ \hat{y} \circ \lambda_{\mathbb{X}} \circ x \in \mathrm{End}_{\mathcal{O}_E}(\overline{\mathbb{Y}}) \xrightarrow{\sim} \mathcal{O}_E.$$

By [Kudla and Rapoport 2011, Lemma 3.9], there is an isomorphism of σ-Hermitian E-spaces

$$\mathrm{Hom}_{\mathcal{O}_E}(\overline{\mathbb{Y}}, \mathbb{X}) \otimes_{\mathcal{O}_E} E \xrightarrow{\sim} C. \tag{2.3.0.1}$$

Therefore we may view elements of C as special quasi-homomorphisms.

2.4. ***Vertex lattices.*** For any $\mathcal{O}_E$-lattice $\Lambda \subset C$, we define the dual lattice $\Lambda^\vee := \{x \in C : \{x, \Lambda\} \subseteq \mathcal{O}_E\}$. It follows from the σ-Hermitian property (2.2.0.5) that we have $(\Lambda^\vee)^\vee = \Lambda$.

A *vertex lattice* is an $\mathcal{O}_E$-lattice $\Lambda \subseteq C$ such that $p\Lambda \subseteq \Lambda^\vee \subseteq \Lambda$. Such lattices correspond to the vertices of the Bruhat–Tits building of the unitary group $\mathrm{U}(C)$. Fix a vertex lattice Λ. The *type* of Λ is defined to be $t_\Lambda := \dim_{\mathbb{F}_{p^2}} \Lambda/\Lambda^\vee$, which is always an *odd* integer such that $1 \le t_\Lambda \le n$ (see [Vollaard 2010, Remark 2.3]).

We define $\Omega_0(\Lambda) := \Lambda/\Lambda^\vee$ and equip it with the perfect σ-Hermitian form

$$(\cdot,\cdot) : \Omega_0(\Lambda) \times \Omega_0(\Lambda) \to \mathbb{F}_{p^2}, \quad (x, y) := p\{\tilde{x}, \tilde{y}\} \bmod p,$$

where $\{\cdot,\cdot\}$ is the Hermitian form on C defined in (2.2.0.3), and $\tilde{x}, \tilde{y} \in \Lambda$ are lifts of x and y.

We define

$$\Omega(\Lambda) := \Omega_0(\Lambda) \otimes_{\mathbb{F}_{p^2}} k.$$

Remark 2.4.1. Our $\Omega_0(\Lambda)$ is the space V in [Vollaard 2010, (2.11)], and our pairing $(\cdot,\cdot)$ differs from the pairing $(\cdot,\cdot)$ defined in [loc. cit.] by a factor of the reduction $\bar{\delta} \in \mathbb{F}_{p^2}^\times$ of δ.

2.5. ***The variety $\mathcal{V}(\Lambda)$.*** Let Λ be a vertex lattice and let $\Omega_0 = \Omega_0(\Lambda)$. Recall from Section 2.4 that Ω_0 is an $\mathbb{F}_{p^2}$-vector space whose dimension is equal to the type $t = t_\Lambda$ of Λ, an odd number. Let $d := (t-1)/2$. We define $\mathcal{V}(\Omega_0)$ to be the closed $\mathbb{F}_{p^2}$-subscheme of the Grassmannian $\mathrm{Gr}_{d+1}(\Omega_0)$ (viewed as a scheme over $\mathbb{F}_{p^2}$) such that for any $\mathbb{F}_{p^2}$-algebra R,

$$\mathcal{V}(\Omega_0)(R) = \{R\text{-module local direct summands } U \subseteq \Omega_0 \otimes_{\mathbb{F}_{p^2}} R : \\ \operatorname{rank} U = d+1 \text{ and } U^\perp \subseteq U\}. \quad (2.5.0.1)$$

Here $U^\perp$ is by definition $\{v \in \Omega_0 \otimes R : (v, u)_R = 0, \forall u \in U\}$, where $(\cdot, \cdot)_R$ is the R-sesquilinear form on $\Omega_0 \otimes R$ obtained from $(\cdot, \cdot)$ by extension of scalars (linearly in the first variable and σ-linearly in the second variable). Then $\mathcal{V}(\Omega_0)$ is a smooth projective $\mathbb{F}_{p^2}$-scheme of dimension d by [Vollaard 2010, Proposition 2.13] and Remark 2.4.1. In fact, $\mathcal{V}(\Omega_0)$ can be identified as a (generalized) Deligne–Lusztig variety, by [Vollaard and Wedhorn 2011, §4.5] (though we will not use this identification in the following).

We write $\mathcal{V}(\Lambda)$ for the base change of $\mathcal{V}(\Omega_0)$ from $\mathbb{F}_{p^2}$ to k.

2.6. ***Structure of the reduced scheme $\mathcal{N}^{\mathrm{red}}$.*** For each vertex lattice $\Lambda \subseteq C$, we define $\mathcal{N}_\Lambda \subseteq \mathcal{N}$ to be the locus where $\rho_X^{-1} \circ \Lambda^\vee \subseteq \operatorname{Hom}(\bar{Y}, X)$, i.e., where the quasi-homomorphisms $\rho^{-1} \circ v$ lift to actual homomorphisms for any $v \in \Lambda^\vee$. Then $\mathcal{N}_\Lambda$ is a closed formal subscheme by [Rapoport and Zink 1996, Proposition 2.9]. By [Vollaard and Wedhorn 2011, §4] we have an isomorphism of k-varieties

$$\mathcal{N}_\Lambda^{\mathrm{red}} \xrightarrow{\sim} \mathcal{V}(\Lambda). \quad (2.6.0.1)$$

2.7. ***Some invariants associated to a k-point of $\mathcal{N}$.*** We follow [Kudla and Rapoport 2011, §2.1].

Let x be a point in $\mathcal{N}(k)$. Then x represents a tuple $(X, \iota, \lambda, \rho)$ over k as recalled in Section 2.1. Via ρ, we view the Dieudonné module of X as a W-lattice M_x in N, which is stable under the operators F and V. The endomorphism structure ι induces an action of $\mathcal{O}_E \otimes_{\mathbb{Z}_p} W \cong W \oplus W$ on M_x, which is equivalent to the structure of a $\mathbb{Z}/2\mathbb{Z}$-grading on M_x (into W-modules). We denote this grading by

$$M_x = \operatorname{gr}_0 M_x \oplus \operatorname{gr}_1 M_x.$$

This grading is compatible with (2.2.0.2) in the sense that

$$\operatorname{gr}_i M_x = M_x \cap N_i, \quad i = 0, 1.$$

Moreover both $\operatorname{gr}_0 M_x$ and $\operatorname{gr}_1 M_x$ are free W-modules of rank n.

Consider the k-vector space $M_{x,k} := M_x \otimes_W k$. It has an induced $\mathbb{Z}/2\mathbb{Z}$-grading, as well as a canonical filtration $\operatorname{Fil}^1(M_{x,k}) \subset M_{x,k}$. Explicitly, $\operatorname{Fil}^1(M_{x,k})$ is the image of $V(M_x) \subseteq M_x$ under the reduction map $M_x \to M_{x,k}$. Define

$$\operatorname{Fil}^1(\operatorname{gr}_i M_{x,k}) := \operatorname{Fil}^1(M_{x,k}) \cap \operatorname{gr}_i M_{x,k}.$$

Then

$$\operatorname{Fil}^1(M_{x,k}) = \operatorname{Fil}^1(\operatorname{gr}_0 M_{x,k}) \oplus \operatorname{Fil}^1(\operatorname{gr}_1 M_{x,k}),$$

and by the signature $(1, n-1)$ condition we know that $\mathrm{Fil}^1(\mathrm{gr}_0\, M_{x,k})$ is a hyperplane and $\mathrm{Fil}^1(\mathrm{gr}_1\, M_{x,k})$ is a line in $\mathrm{gr}_0\, M_{x,k}$ and $\mathrm{gr}_1\, M_{x,k}$, respectively.

The symplectic form $\langle\cdot,\cdot\rangle$ on N takes values in W on M_x, and hence induces a symplectic form on $M_{x,k}$ by reduction. The latter restricts to a k-bilinear nondegenerate pairing

$$\mathrm{gr}_0\, M_{x,k} \times \mathrm{gr}_1\, M_{x,k} \to k.$$

Under the above pairing, the spaces $\mathrm{Fil}^1(\mathrm{gr}_0\, M_{x,k})$ and $\mathrm{Fil}^1(\mathrm{gr}_1\, M_{x,k})$ are annihilators of each other. Equivalently, $\mathrm{Fil}^1(M_{x,k})$ is a totally isotropic subspace of $M_{x,k}$.

2.8. ***Description of k-points by special lattices.*** For a W-lattice A in N_0, we define its dual lattice to be $A^\vee := \{x \in N_0 : \{x, A\} \subseteq W\}$. If Λ is an $\mathcal{O}_E$-lattice in C, then we have $(\Lambda_W)^\vee = (\Lambda^\vee)_W$. In the following we denote both of them by $\Lambda_W^\vee$.

Definition 2.8.1. A *special lattice* is a W-lattice A in N_0 such that

$$A^\vee \subseteq A \subseteq p^{-1}A^\vee$$

and such that $A/A^\vee$ is a one-dimensional k-vector space.

Remark 2.8.2. The apparent difference between the above definition and the condition in [Vollaard 2010, Proposition 1.10] (for $i = 0$) is caused by the fact that we have normalized the pairing $\{\cdot,\cdot\}$ on N_0 differently from [loc. cit.], using an extra factor $(p\delta)^{-1}$ (see (2.2.0.3)). Our normalization is the same as that in [RTZ].

Recall the following result.

Proposition 2.8.3 [Vollaard 2010, Proposition 1.10]. *There is a bijection from $\mathcal{N}(k)$ to the set of special lattices, sending a point x to $\mathrm{gr}_0\, M_x$ considered in Section 2.7.* □

Remark 2.8.4. Let $x \in \mathcal{N}(k)$ and let A be the special lattice associated to it by Proposition 2.8.3. Let Λ be a vertex lattice. Then $x \in \mathcal{N}_\Lambda(k)$ if and only if $A \subseteq \Lambda_W$, if and only if $\Lambda_W^\vee \subseteq A^\vee$. (See also Remark 3.1.5 below.)

2.9. ***Filtrations.*** We introduce the following notation:

Definition 2.9.1. Let A be a special lattice. Write $A_k := A \otimes_W k$. Let $x \in \mathcal{N}(k)$ correspond to A under Proposition 2.8.3. Thus $A_k = \mathrm{gr}_0\, M_{x,k}$. Define $\mathrm{Fil}^1(A_k) := \mathrm{Fil}^1(\mathrm{gr}_0\, M_{x,k})$ (see Section 2.7). It is a hyperplane in A_k.

Lemma 2.9.2. *Let A be a special lattice. Then $\Phi^{-1}(A^\vee)$ is contained inside A, and its image in A_k is equal to* $\mathrm{Fil}^1(A_k)$.

Proof. Let A correspond to $x \in \mathcal{N}(k)$ under Proposition 2.8.3. Then F and V both preserve the W-lattice M_x in N (see Section 2.7). By definition, $\mathrm{Fil}^1(M_{x,k})$ is the image of $V(M_x) \subseteq M_x$ under the reduction map $M_x \to M_{x,k}$. Since the operator V

is of degree 1 with respect to the $\mathbb{Z}/2\mathbb{Z}$-grading, we see that $\mathrm{Fil}^1(A_k)$ is the image of $V(\mathrm{gr}_1 M_x) \subseteq A$ under $A \to A_k$. It suffices to prove that

$$\Phi^{-1}(A^\vee) = V(\mathrm{gr}_1 M_x). \tag{2.9.2.1}$$

By the proof of [Vollaard 2010, Proposition 1.10], we have $\mathrm{gr}_1 M_x = F^{-1}A^\vee$. (Note that because of the difference of normalizations as discussed in Remark 2.8.2, what is denoted by $A^\vee$ here is denoted by $pA^\vee$ in [loc. cit.]. Also note that the integer i appearing [loc. cit.] is 0 in our case.) Therefore $V(\mathrm{gr}_1 M_x) = V(F^{-1}A^\vee)$. But $VF^{-1} = (V^{-1}F)^{-1} = \Phi^{-1}$ because $VF = FV = p$. Thus (2.9.2.1) holds as desired. $\square$

2.10. ***Special cycles.*** Let $\boldsymbol{v}$ be an arbitrary subset of C. We define the *special cycle* $\mathcal{Z}(\boldsymbol{v}) \subseteq \mathcal{N}$ to be the locus where $\rho^{-1} \circ v \in \mathrm{Hom}(\overline{Y}, X)$ for all $v \in \boldsymbol{v}$, i.e., all the quasi-homomorphisms $\rho^{-1} \circ v$ lift to actual homomorphisms. Note that $\mathcal{Z}(\boldsymbol{v})$ only depends on the $\mathcal{O}_E$-submodule $L(\boldsymbol{v})$ spanned by $\boldsymbol{v}$ in C, and we have $\mathcal{Z}(\boldsymbol{v}) = \mathcal{Z}(L(\boldsymbol{v}))$.

We say $\boldsymbol{v}$ is *minuscule* if $L(\boldsymbol{v})$ is an $\mathcal{O}_E$-lattice in C satisfying $pL(\boldsymbol{v})^\vee \subseteq L(\boldsymbol{v}) \subseteq L(\boldsymbol{v})^\vee$, or equivalently, if $L(\boldsymbol{v})$ is the dual of a vertex lattice. When this is the case we have $\mathcal{Z}(\boldsymbol{v}) = \mathcal{N}_{L(\boldsymbol{v})^\vee}$ by definition.

2.11. ***The intersection problem.*** Let $C^\flat$ be the analogue for $\mathcal{M}$ of the Hermitian space C. Then $C \cong C^\flat \oplus Eu$ for some vector denoted by u which is of norm 1 and orthogonal to $C^\flat$. We have a closed immersion

$$\delta : \mathcal{M} \to \mathcal{N},$$

sending $(X, \iota, \lambda, \rho)$ to $(X \times \overline{Y}, \iota \times \iota_{\overline{Y}}, \lambda \times \lambda_{\overline{Y}}, \rho \times \mathrm{id})$. We have $\delta(\mathcal{M}) = \mathcal{Z}(u)$. The closed immersion δ induces a closed immersion of formal schemes

$$(\mathrm{id}, \delta) : \mathcal{M} \to \mathcal{M} \times_W \mathcal{N}.$$

Denote by Δ the image of (id, δ), which we call the (local) *GGP cycle*. For any $g \in J(\mathbb{Q}_p)$, we obtain a formal subscheme

$$(\mathrm{id} \times g)\Delta \subseteq \mathcal{M} \times_W \mathcal{N},$$

via the action of g on $\mathcal{N}$. Let $g \in J(\mathbb{Q}_p)$ and let $\mathcal{N}^g \subseteq \mathcal{N}$ be the fixed locus of g. Then by definition we have

$$\Delta \cap (\mathrm{id} \times g)\Delta \cong \delta(\mathcal{M}) \cap \mathcal{N}^g.$$

Our goal is to compute the arithmetic intersection number

$$\langle \Delta, (\mathrm{id} \times g)\Delta \rangle,$$

when g is *regular semisimple* and *minuscule* (as defined in the introduction). Notice that $g \in J(\mathbb{Q}_p)$ is regular semisimple if and only if $\boldsymbol{v}(g) := (u, gu, \dots, g^{n-1}u)$ is an E-basis of C. Also notice that a regular semisimple element g is minuscule if and only if $\boldsymbol{v}(g)$ is minuscule in the sense of Section 2.10.

3. Reducedness of minuscule special cycles

3.1. *Local structure of special cycles.*

Definition 3.1.1. Let $\mathscr{C}$ be the following category:

- Objects in $\mathscr{C}$ are triples $(\mathcal{O}, \mathcal{O} \to k, \delta)$, where $\mathcal{O}$ is a local Artinian W-algebra, $\mathcal{O} \to k$ is a W-algebra map, and δ is a nilpotent divided power structure on $\ker(\mathcal{O} \to k)$ (see [Berthelot and Ogus 1978, Definitions 3.1, 3.27]).
- Morphisms in $\mathscr{C}$ are W-algebra maps that are compatible with the structure maps to k and the divided power structures.

3.1.2. Let $x \in \mathcal{N}(k)$ correspond to a special lattice A under Proposition 2.8.3. Let $\mathcal{O} \in \mathscr{C}$. By a *hyperplane* in $A_{\mathcal{O}} := A \otimes_W \mathcal{O}$ we mean a free direct summand of $A_{\mathcal{O}}$ of rank $n-1$. We define the $\mathbb{Z}/2\mathbb{Z}$-grading on $M_{x,\mathcal{O}} := M_x \otimes_W \mathcal{O}$ by linearly extending that on M_x (see Section 2.7). Denote by $\widehat{\mathcal{N}}_x$ the completion of $\mathcal{N}$ at x. For any $\tilde{x} \in \widehat{\mathcal{N}}_x(\mathcal{O})$, we have a unitary p-divisible group of signature $(1, n-1)$ over $\mathcal{O}$ deforming that over k defined by x. By Grothendieck–Messing theory, we obtain the Hodge filtration $\mathrm{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} \subseteq M_{x,\mathcal{O}}$. Define $f_{\mathcal{O}}(\tilde{x})$ to be the intersection

$$\mathrm{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} \cap \mathrm{gr}_0 M_{x,\mathcal{O}}$$

inside $M_{x,\mathcal{O}}$. By the signature $(1, n-1)$ condition, $f_{\mathcal{O}}(\tilde{x})$ is a hyperplane in $A_{\mathcal{O}}$. It also lifts $\mathrm{Fil}^1 A_k$ (see Definition 2.9.1) by construction. Thus we have defined a map

$$f_{\mathcal{O}} : \widehat{\mathcal{N}}_x(\mathcal{O}) \xrightarrow{\sim} \{\text{hyperplanes in } A_{\mathcal{O}} \text{ lifting } \mathrm{Fil}^1 A_k\}. \tag{3.1.2.1}$$

By construction, $f_{\mathcal{O}}$ is functorial in $\mathcal{O}$ in the sense that the collection $(f_{\mathcal{O}})_{\mathcal{O}\in\mathscr{C}}$ is a natural transformation between two set-valued functors on $\mathscr{C}$. Here we are viewing the right hand side of (3.1.2.1) as a functor in $\mathcal{O}$ using the base change maps.

The following result is the analogue of [Li and Zhu 2017, Theorem 4.1.7]. As a direct consequence of the PEL moduli problem, it should be well known to the experts and is essentially proved in [Kudla and Rapoport 2011, Proposition 3.5].

Theorem 3.1.3. *Keep the notations in Section 3.1.2*:

(1) *The natural transformation $(f_{\mathcal{O}})_{\mathcal{O}\in\mathscr{C}}$ is an isomorphism.*

(2) *Let $\boldsymbol{v}$ be a subset of C. If $x \in \mathcal{Z}(\boldsymbol{v})(k)$, then $\boldsymbol{v} \subseteq A$. Suppose $x \in \mathcal{Z}(\boldsymbol{v})(k)$. Then for any $\mathcal{O} \in \mathscr{C}$ the map $f_{\mathcal{O}}$ induces a bijection*

$$\widehat{\mathcal{Z}(\boldsymbol{v})}_x(\mathcal{O}) \xrightarrow{\sim} \{\textit{hyperplanes in } A_{\mathcal{O}} \textit{ lifting } \operatorname{Fil}^1 A_k \textit{ and containing the image of } \boldsymbol{v} \textit{ in } A_{\mathcal{O}}\}.$$

Proof. (1) We need to check that for all $\mathcal{O} \in \mathscr{C}$ the map $f_{\mathcal{O}}$ is a bijection. Let $\tilde{x} \in \widehat{\mathcal{N}}_x(\mathcal{O})$. This represents a deformation over $\mathcal{O}$ of the p-divisible group at x. Similarly to the situation in Section 2.7, the compatibility with the endomorphism structure implies that

$$\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} = \bigoplus_{i=0}^{1} \operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} \cap \operatorname{gr}_i M_{x,\mathcal{O}}.$$

By the compatibility with the polarization, we know that $\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}}$ is totally isotropic under the symplectic form on $M_{x,\mathcal{O}}$. It follows that the two modules $\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} \cap \operatorname{gr}_1 M_{x,\mathcal{O}}$ and $\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}} \cap \operatorname{gr}_0 M_{x,\mathcal{O}}$ are annihilators of each other if we identify $\operatorname{gr}_1 M_{x,\mathcal{O}}$ as the $\mathcal{O}$-linear dual of $\operatorname{gr}_0 M_{x,\mathcal{O}}$ using the symplectic form on $M_{x,\mathcal{O}}$. Therefore, $\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}}$ can be recovered from $f_{\mathcal{O}}(\tilde{x})$. This together with Grothendieck–Messing theory proves the injectivity of $f_{\mathcal{O}}$. The surjectivity of $f_{\mathcal{O}}$ also follows from Grothendieck–Messing theory and the above way of reconstructing $\operatorname{Fil}^1_{\tilde{x}} M_{x,\mathcal{O}}$ from its intersection with $\operatorname{gr}_0 M_{x,\mathcal{O}}$. Note that the unitary p-divisible groups reconstructed in this way do satisfy the signature condition because we have started with *hyperplanes* in $A_{\mathcal{O}}$.

(2) The statements follow from the proof of [Kudla and Rapoport 2011, Proposition 3.5] and the definition of (2.3.0.1) in [Kudla and Rapoport 2011, Lemma 3.9]. We briefly recall the arguments here. If $\phi \in \operatorname{Hom}_{\mathcal{O}_E}(\overline{\mathbb{Y}}, \mathbb{X}) \otimes_{\mathcal{O}_E} E$ is a special quasi-homomorphism, the element $v \in C$ corresponding to ϕ under (2.3.0.1) is by definition the projection to N_0 of $\phi_*(\bar{1}_0) \in N$, where ϕ_* is the map $\mathbb{D}(\overline{\mathbb{Y}}) \otimes_W K \to \mathbb{D}(\mathbb{X}) \otimes_W K = N$ induced by ϕ, and $\bar{1}_0$ is a certain fixed element in $\mathbb{D}(\overline{\mathbb{Y}})$. In fact, $\bar{1}_0$ is chosen such that

- $W\bar{1}_0 = \operatorname{gr}_0 \mathbb{D}(\overline{\mathbb{Y}})$, where the grading is with respect to the $\mathcal{O}_E$-action on $\overline{\mathbb{Y}}$,
- $W\bar{1}_0 = \operatorname{Fil}^1_{\overline{Y}} \mathbb{D}(\overline{\mathbb{Y}})$, the Hodge filtration for the deformation $\overline{Y}$ of $\overline{\mathbb{Y}}$ over W.

In particular v and ϕ are related by the formula $v = \phi_*(\bar{1}_0)$, as the projection to N_0 is not needed.

From now on we assume without loss of generality that $\boldsymbol{v} = \{v\}$, with v corresponding to ϕ as in the above paragraph. If $x \in \mathcal{Z}(v)(k)$, then ϕ_* has to map $\mathbb{D}(\overline{\mathbb{Y}})$ into M_x, so $v \in M_x$. Since ϕ_* is compatible with the $\mathbb{Z}/2\mathbb{Z}$-gradings, we further have $v \in A$. We have shown that if $x \in \mathcal{Z}(v)(k)$, then $v \in A$.

Now suppose $x \in \mathcal{Z}(\boldsymbol{v})(k)$. Let $\mathcal{O} \in \mathscr{C}$. Write $v_{\mathcal{O}} := v \otimes 1 \in A_{\mathcal{O}} \subset M_{x,\mathcal{O}}$. For all $\tilde{x} \in \widehat{\mathcal{N}}_x(\mathcal{O})$, by Grothendieck–Messing theory we know that $\tilde{x} \in \widehat{\mathcal{Z}(v)}_x(\mathcal{O})$ if and only if the base change of ϕ_* to $\mathcal{O}$ (still denoted by ϕ_*) preserves the Hodge filtrations, i.e.,

$$\phi_*(\mathrm{Fil}^1_{\bar{Y}}\, \mathbb{D}(\overline{\mathbb{Y}})) \subseteq \mathrm{Fil}^1_{\tilde{x}}\, M_{x,\mathcal{O}}.$$

Since $W\bar{1}_0 = \mathrm{Fil}^1_{\bar{Y}}\, \mathbb{D}(\overline{\mathbb{Y}})$, this last condition is equivalent to $v_{\mathcal{O}} \in \mathrm{Fil}^1_{\tilde{x}}\, M_{x,\mathcal{O}}$. Again, because ϕ_* is compatible with the $\mathbb{Z}/2\mathbb{Z}$-gradings, the last condition is equivalent to $v_{\mathcal{O}} \in f_{\mathcal{O}}(\tilde{x})$. In conclusion, we have shown that $\tilde{x} \in \widehat{\mathcal{N}}_x(\mathcal{O})$ is in $\widehat{\mathcal{Z}(v)}_x(\mathcal{O})$ if and only if $v_{\mathcal{O}} \in f_{\mathcal{O}}(\tilde{x})$, as desired. □

Corollary 3.1.4. *Let $x \in \mathcal{N}(k)$ correspond to the special lattice A. Let $\boldsymbol{v}$ be a subset of C. Then $x \in \mathcal{Z}(\boldsymbol{v})(k)$ if and only if $\boldsymbol{v} \subseteq A^\vee$.*

Proof. By part (2) of Theorem 3.1.3 applied to $\mathcal{O} = k$, we see that $x \in \mathcal{Z}(\boldsymbol{v})(k)$ if and only if $\boldsymbol{v} \subseteq A$ and the image of $\boldsymbol{v}$ in A_k is contained in $\mathrm{Fil}^1(A_k)$. The corollary follows from Lemma 2.9.2 and the Φ-invariance of $\boldsymbol{v}$. □

Remark 3.1.5. Note that Remark 2.8.4 is a special case of Corollary 3.1.4.

3.2. *Proof of the reducedness.*

Corollary 3.2.1. *Let Λ be an $\mathcal{O}_E$-lattice in C with $p^i\Lambda \subseteq \Lambda^\vee \subseteq \Lambda$ for some $i \in \mathbb{Z}_{\geq 1}$. Then the special cycle $\mathcal{Z}(\Lambda^\vee)$ defined by $\Lambda^\vee$ has no (W/p^{i+1})-points. In particular, taking $i = 1$ we see that $\mathcal{N}_\Lambda(W/p^2) = \varnothing$ for any vertex lattice Λ.*

Proof. Let $\mathcal{O} = W/p^{i+1}$, equipped with the reduction map $W/p^{i+1} \to k$ and the natural divided power structure on the kernel $p\mathcal{O}$. Then $\mathcal{O} \in \mathscr{C}$. Assume $\mathcal{Z}(\Lambda^\vee)$ has an $\mathcal{O}$-point $\tilde{x}$ reducing to a k-point x. Let A be the special lattice corresponding to x (see Section 2.8). By Theorem 3.1.3, there exists a hyperplane P in $A_{\mathcal{O}}$ lifting $\mathrm{Fil}^1(A_k)$, such that $P \supseteq \Lambda^\vee \otimes_{\mathcal{O}_E} \mathcal{O}$. Since P is a hyperplane in $A_{\mathcal{O}}$, there exists an element $l \in \mathrm{Hom}_{\mathcal{O}}(A_{\mathcal{O}}, \mathcal{O})$ such that

$$l(P) = 0 \quad \text{and} \quad l(A_{\mathcal{O}}) = \mathcal{O}. \tag{3.2.1.1}$$

We may find an element $\tilde{l} \in A^\vee \subseteq N_0$ to represent l, in the sense that for all $a \otimes 1 \in A_{\mathcal{O}}$ with $a \in A$, we have

$$l(a \otimes 1) = \text{the image of } \{a, \tilde{l}\} \text{ under } W \to \mathcal{O}.$$

Since $l(\Lambda^\vee \otimes \mathcal{O}) \subseteq l(P) = 0$, we know that $\{v, \tilde{l}\} \in p^{i+1}W$ for all $v \in \Lambda^\vee$. Since $\Lambda^\vee \subseteq C = N_0^\Phi$, applying (2.2.0.4) we see that $\{\tilde{l}, v\} \in p^{i+1}W$ for all $v \in \Lambda^\vee$. Therefore

$$p^{-i-1}\tilde{l} \in (\Lambda_W^\vee)^\vee = \Lambda_W,$$

and thus $\tilde{l} \in p^{i+1}\Lambda_W$, which is contained in $p\Lambda_W^\vee$ by hypothesis. Since Λ is Φ-invariant, we also have $\Phi(\tilde{l}) \in p\Lambda_W^\vee$. But $\Lambda_W^\vee \subseteq A^\vee$ by Corollary 3.1.4, so $\Phi(\tilde{l}) \in pA^\vee$. It follows that for all $a \in A$, we have $\{\Phi(\tilde{l}), a\} \in pW$, and therefore

$$\{a, \tilde{l}\} \overset{(2.2.0.4)}{=} \sigma^{-1}(\{\Phi(\tilde{l}), a\}) \in pW,$$

contradicting the second condition in (3.2.1.1). □

Corollary 3.2.2. *Let Λ be a vertex lattice of type t and let $x \in \mathcal{N}_\Lambda(k)$. Then the tangent space $\mathcal{T}_x\mathcal{N}_{\Lambda,k}$ to $\mathcal{N}_{\Lambda,k}$ at x, where $\mathcal{N}_{\Lambda,k}$ is the special fiber* (*i.e., base change to k*) *of $\mathcal{N}_\Lambda$, is of k-dimension* $(t-1)/2$.

Proof. This can be deduced from Theorem 3.1.3 elementarily, in the same way as in [Li and Zhu 2017, §4.2]. Here we provide a shorter proof. Firstly we make an easy observation. Denote by $\mathscr{C}_k$ the full subcategory of $\mathscr{C}$ consisting of characteristic p objects. Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be two formal schemes over k. For $i = 1, 2$ fix $y_i \in \mathcal{W}_i(k)$ and define the set-valued functor $\mathcal{F}_i$ on $\mathscr{C}_k$ sending $\mathcal{O}$ to the set of $\mathcal{O}$-points of $\mathcal{W}_i$ which induce y_i under the structure map $\mathcal{O} \to k$. Assume $\mathcal{F}_1 \cong \mathcal{F}_2$. Then the tangent spaces $\mathcal{T}_{x_i}\mathcal{W}_i$ are isomorphic. In fact, this observation is a direct consequence of the definition of the vector space structure on the tangent spaces from the point of view of functor of points, as recalled in the proof of [Li and Zhu 2017, Lemma 4.2.6] for instance.

Denote by B the k-subspace of A_k spanned by the image of $\Lambda^\vee$ in A_k. Consider the Grassmannian $\mathrm{Gr}_{n-1}(A_k)$ parametrizing hyperplanes in the n-dimensional k-vector space A_k. Let $\mathcal{W}_1$ be the subvariety of $\mathrm{Gr}_{n-1}(A_k)$ defined by the condition that the hyperplane should contain B, and let $y_1 \in \mathcal{W}_1(k)$ corresponding to $\mathrm{Fil}^1(A_k) \subseteq A_k$. Let $\mathcal{W}_2 := \mathcal{N}_{\Lambda,k}$ and $y_2 := x$. By Theorem 3.1.3, the assumption on $(\mathcal{W}_i, y_i)$, $i = 1, 2$ in the previous paragraph is satisfied. Hence it suffices to compute the dimension of $\mathcal{T}_{y_1}\mathcal{W}_1$. Note that $\mathcal{W}_1$ is itself a Grassmannian, parametrizing hyperplanes in A_k/B. The proof is finished once we know that A_k/B has k-dimension $(t+1)/2$. But this is true by the (σ-linear) duality between the k-vector spaces $A_k/B = A/\Lambda_W^\vee$ and $\Lambda_W/A^\vee$ under the σ-sesquilinear form on $\Omega(\Lambda)$ obtained by extension of scalars from the σ-Hermitian form $(\cdot\,,\cdot)$ on $\Omega_0(\Lambda)$ (see Section 2.4) and the fact that $A/A^\vee$ is a 1-dimensional k-vector space (see Definition 2.8.1). □

In the following corollary we reprove [RTZ, Theorems 9.4 and 10.1].

Corollary 3.2.3. *Let Λ be a vertex lattice. Then $\mathcal{N}_\Lambda = \mathcal{N}_{\Lambda,k} = \mathcal{N}_\Lambda^{\mathrm{red}}$ and it is regular.*

Proof. Let t be the type of Λ. Recall from Section 2.6 that $\mathcal{N}_\Lambda^{\mathrm{red}}$ is a smooth k-scheme of dimension $(t-1)/2$. By Corollary 3.2.2, all the tangent spaces of $\mathcal{N}_{\Lambda,k}$ have k-dimension $(t-1)/2$, and so $\mathcal{N}_{\Lambda,k}$ is regular. In particular $\mathcal{N}_{\Lambda,k}$ is reduced, namely $\mathcal{N}_{\Lambda,k} = \mathcal{N}_\Lambda^{\mathrm{red}}$. Knowing that $\mathcal{N}_{\Lambda,k}$ is regular, and that $\mathcal{N}_\Lambda$ has no (W/p^2)-points (Corollary 3.2.1), it follows that $\mathcal{N}_\Lambda = \mathcal{N}_{\Lambda,k}$ by the general criterion [RTZ, Lemma 10.3]. □

4. The intersection length formula

4.1. ***The arithmetic intersection as a fixed point scheme.*** Fix a regular semisimple and minuscule element $g \in J(\mathbb{Q}_p)$. Let $L := L(\boldsymbol{v}(g))$ and $\Lambda := L^\vee$. They are both $\mathcal{O}_E$-lattices in C. Recall from the end of Section 2.10 that Λ is a vertex lattice and $\mathcal{Z}(L) = \mathcal{N}_\Lambda$. From now on we assume $\mathcal{N}^g(k) \neq \varnothing$. As shown in [RTZ, §5], this assumption implies that both L and Λ are g-cyclic and stable under g. In particular, the natural action of g on $\mathcal{N}$ stabilizes $\mathcal{N}_\Lambda$.

Let $\Omega_0 = \Omega_0(\Lambda)$ and $\Omega = \Omega(\Lambda)$. Let $t = t_\Lambda$ and $d = (t-1)/2$ as in Section 2.5. Let $\bar{g} \in \mathrm{GL}(\Omega_0)(\mathbb{F}_{p^2})$ be the induced action of g on Ω_0. Then $\bar{g}$ preserves the Hermitian form $(\cdot\,,\cdot)$ on Ω_0 and hence acts on $\mathcal{V}(\Lambda)$. It is clear from the definition of the isomorphism (2.6.0.1) given in [Vollaard and Wedhorn 2011, §4] that it is equivariant for the actions of g and $\bar{g}$ on the two sides.

Remark 4.1.1. Since Λ and $\Lambda^\vee$ are g-cyclic, the linear operator $\bar{g} \in \mathrm{GL}(\Omega_0)(\mathbb{F}_{p^2})$ has equal minimal polynomial and characteristic polynomial. Equivalently, in the Jordan normal form of $\bar{g}$ (over k) there is a unique Jordan block associated to any eigenvalue.

Proposition 4.1.2. *$\delta(\mathcal{M}) \cap \mathcal{N}^g$ is a scheme of characteristic p (i.e., a k-scheme) isomorphic to $\mathcal{V}(\Lambda)^{\bar{g}}$.*

Proof. Recall from Section 2.11 that $\delta(\mathcal{M}) = \mathcal{Z}(u)$. Since the $\mathcal{O}_E$-module L is generated by $u, gu, \cdots, g^{n-1}u$ and stable under g, we have $\delta(\mathcal{M}) \cap \mathcal{N}^g = \mathcal{Z}(L)^g = \mathcal{N}_\Lambda^g$. By Corollary 3.2.3, we know that $\mathcal{N}_\Lambda^g = (\mathcal{N}_\Lambda^{\mathrm{red}})^g$. But the latter is isomorphic to the characteristic p scheme $\mathcal{V}(\Lambda)^{\bar{g}}$ under (2.6.0.1). □

4.2. ***Study of $\mathcal{V}(\Lambda)^{\bar{g}}$.*** We start with an alternative moduli interpretation of $\mathcal{V}(\Omega_0)$. The idea is to rewrite (in Lemma 4.2.2) the procedure of taking the complement $U \mapsto U^\perp$ with respect to the Hermitian form, in terms of taking the complement with respect to some quadratic form and taking a Frobenius. The alternative moduli interpretation is given in Corollary 4.2.3 below.

Let Θ_0 be a t-dimensional nondegenerate quadratic space over $\mathbb{F}_p$. Let $[\cdot\,,\cdot] : \Theta_0 \times \Theta_0 \to \mathbb{F}_p$ be the associated bilinear form. Since there is a unique isomorphism class of nondegenerate σ-Hermitian spaces over $\mathbb{F}_{p^2}$, we may assume that $\Omega_0 = \Theta_0 \otimes_{\mathbb{F}_p} \mathbb{F}_{p^2}$ and that the σ-Hermitian form $(\cdot\,,\cdot)$ (see Section 2.5) is obtained by extension of scalars (linearly in the first variable and σ-linearly in the second variable) from $[\cdot\,,\cdot]$.

Definition 4.2.1. Let R be an $\mathbb{F}_p$-algebra. We define $[\cdot\,,\cdot]_R$ to be the R-bilinear form on $\Theta_0 \otimes_{\mathbb{F}_p} R$ obtained from $[\cdot\,,\cdot]$ by extension of scalars. For any R-submodule $\mathcal{L} \subset \Theta_0 \otimes_{\mathbb{F}_p} R$, define

$$\mathcal{L}^{\mathrm{lin}\perp} := \{v \in \Theta_0 \otimes_{\mathbb{F}_p} R : [v, l]_R = 0, \forall l \in \mathcal{L}\}.$$

Define $\sigma_*(\mathcal{L})$ to be the R-module generated by the image of $\mathcal{L}$ under the map

$$\sigma : \Theta_0 \otimes_{\mathbb{F}_p} R \to \Theta_0 \otimes_{\mathbb{F}_p} R, \quad v \otimes r \mapsto v \otimes r^p.$$

Let R be an $\mathbb{F}_{p^2}$-algebra. Let U be an R-submodule of $\Omega_0 \otimes_{\mathbb{F}_{p^2}} R$. Since $\Omega_0 \otimes_{\mathbb{F}_{p^2}} R = \Theta_0 \otimes_{\mathbb{F}_p} R$, we may view U as an R-submodule of the latter and define $\sigma_*(U)$ as in Definition 4.2.1.

Lemma 4.2.2. *We have $U^\perp = (\sigma_*(U))^{\mathrm{lin}\perp}$.*

Proof. Consider two arbitrary elements

$$x = \sum_j u_j \otimes r_j \quad \text{and} \quad y = \sum_k v_k \otimes s_k$$

of $\Theta_0 \otimes_{\mathbb{F}_p} R$. We have

$$(y, x)_R = \sum_{j,k} s_k r_j^p \cdot [v_k, u_j]_R = \sum_{j,k} s_k r_j^p \cdot [u_j, v_k]_R = [\sigma(x), y]_R.$$

Hence for $y \in \Theta_0 \otimes_{\mathbb{F}_p} R$, we have $y \in U^\perp$ if and only if $(y, x)_R = 0$ for all $x \in U$, if and only if $[\sigma(x), y]_R = 0$ for all $x \in U$, if and only if $y \in (\sigma_*(U))^{\mathrm{lin}\perp}$. □

Corollary 4.2.3. *For any $\mathbb{F}_{p^2}$-algebra R, the set $\mathcal{V}(\Omega_0)(R)$ is equal to the set of R-submodules U of*

$$\Omega_0 \otimes_{\mathbb{F}_{p^2}} R = \Theta_0 \otimes_{\mathbb{F}_p} R,$$

such that U is an R-module local direct summand of rank $d+1$, satisfying

$$(\sigma_*(U))^{\mathrm{lin}\perp} \subseteq U.$$

Proof. This is a direct consequence of (2.5.0.1) and Lemma 4.2.2. □

In the following we denote $\mathcal{V}(\Lambda)$ by $\mathcal{V}$ for simplicity, where Λ is always fixed as in the beginning of Section 4.1. Denote $\Theta := \Theta_0 \otimes_{\mathbb{F}_p} k$. Fix a point $x_0 \in \mathcal{V}^{\bar{g}}(k)$. Let U_0 correspond to x_0 under (2.5.0.1) or Corollary 4.2.3. Define

$$\mathcal{L}_{d+1} := U_0 \quad \text{and} \quad \mathcal{L}_d := (\sigma_*(U_0))^{\mathrm{lin}\perp} \overset{\text{Lemma 4.2.2}}{=} U_0^\perp.$$

They are subspaces of Θ stable under $\bar{g}$, of k-dimensions $d+1$ and d respectively.

Definition 4.2.4. Define $\mathcal{I} := \mathbb{P}(\Theta/\mathcal{L}_d)$, a projective space of dimension d over k.

Then $\mathcal{L}_{d+1}$ defines an element in $\mathcal{I}(k)$, which we still denote by x_0 by abuse of notation. We have a natural action of $\bar{g}$ on $\mathcal{I}$ that fixes x_0. Let $\mathcal{R}_p$ and $\mathcal{S}_p$ be the quotient of the local ring of $\mathcal{I}^{\bar{g}}$ and of $\mathcal{V}^{\bar{g}}$ at x_0 divided by the p-th power of its maximal ideal, respectively.

Lemma 4.2.5. *There is a k-algebra isomorphism $\mathcal{R}_p \cong \mathcal{S}_p$.*

Proof. The proof is based on exactly the same idea as [Li and Zhu 2017, Lemma 5.2.9]. Let $\tilde{\mathcal{R}}_p$ and $\tilde{\mathcal{S}}_p$ be the quotient of the local ring of $\mathcal{I}$ and of $\mathcal{V}$ at x_0 divided by the p-th power of its maximal ideal, respectively. Let R be an arbitrary local k-algebra with residue field k such that the p-th power of its maximal ideal is zero. Then by Lemma 4.2.2, the R-points of $\mathcal{V}$ lifting x_0 classify R-module local direct summands U of $\Theta \otimes_k R$ of rank $d+1$ that lift $\mathcal{L}_{d+1}$, and such that

$$U \supseteq (\sigma_*(U))^{\mathrm{lin}\perp}.$$

But by the assumption that the p-th power of the maximal ideal of R is zero, we have

$$\sigma_{R,*}(U) = (\sigma_{k,*}(\mathcal{L}_{d+1})) \otimes_k R,$$

where we have written σ_R and σ_k to distinguish between the Frobenius on R and on k. Therefore

$$(\sigma_{R,*}(U))^{\mathrm{lin}\perp} = ((\sigma_{k,*}(\mathcal{L}_{d+1})) \otimes_k R)^{\mathrm{lin}\perp} = (\sigma_{k,*}(\mathcal{L}_{d+1}))^{\mathrm{lin}\perp} \otimes_k R = \mathcal{L}_d \otimes_k R.$$

Thus we see that the set of R-points of $\mathcal{V}$ lifting x_0 is in canonical bijection with the set of R-points of $\mathcal{I}$ lifting x_0. We thus obtain a canonical $\tilde{\mathcal{R}}_p$-point of $\mathcal{V}$ lifting $x_0 \in \mathcal{V}(k)$, and a canonical $\tilde{\mathcal{S}}_p$-point of $\mathcal{I}$ lifting $x_0 \in \mathcal{I}(k)$. These two points induce maps $\tilde{\mathcal{S}}_p \to \tilde{\mathcal{R}}_p$ and $\tilde{\mathcal{R}}_p \to \tilde{\mathcal{S}}_p$ respectively. From the moduli interpretation of these two maps we see that they are k-algebra homomorphisms inverse to each other and equivariant with respect to the actions of $\bar{g}$ on both sides. Note that $\mathcal{S}_p$ and $\mathcal{R}_p$ are the quotients of $\tilde{\mathcal{S}}_p$ and $\tilde{\mathcal{R}}_p$ by the augmentation ideal for the $\bar{g}$-action, respectively. It follows that $\mathcal{R}_p \cong \mathcal{S}_p$. □

4.3. *Study of $\mathcal{I}^{\bar{g}}$.*

Definition 4.3.1. Let λ be the eigenvalue of $\bar{g}$ on the 1-dimensional k-vector space $\mathcal{L}_{d+1}/\mathcal{L}_d = U_0/U_0^{\perp}$, and let c be the size of the unique (see Remark 4.1.1) Jordan block of $\bar{g}|_{\mathcal{L}_{d+1}}$ associated to λ. Notice our c is denoted by $c+1$ in [RTZ, §9].

Remark 4.3.2. By the discussion before [Rapoport et al. 2013, Proposition 9.1], c is the size of the unique Jordan block associated to λ of $\bar{g}$ on $\Theta/\mathcal{L}_d = \Omega/U_0^{\perp}$, and is also equal to the quantity $\frac{1}{2}(m(Q(T))+1)$ introduced in the introduction.

Proposition 4.3.3. *The local ring $\mathcal{O}_{\mathcal{I}^{\bar{g}},x_0}$ of $\mathcal{I}^{\bar{g}}$ at x_0 is isomorphic to $k[X]/X^c$ as a k-algebra.*

Proof. By Remark 4.3.2 and Definition 4.2.4, the proposition is a consequence of the following general lemma applied to $\mathcal{L} = \Theta/\mathcal{L}_d$ and $h = \bar{g}$. □

Lemma 4.3.4. *Let $\mathcal{L}$ be a k-vector space of dimension $d+1$. Let $\mathbb{P}(\mathcal{L}) = \mathbb{P}^d$ be the associated projective space. Let $x_0 \in \mathbb{P}(\mathcal{L})(k)$, represented by a vector $\ell \in \mathcal{L}$. Let $h \in \mathrm{GL}(\mathcal{L})(k) = \mathrm{GL}_{d+1}(k)$. Assume that*:

(1) *The natural action of h on $\mathbb{P}(\mathcal{L})$ fixes x_0. Denote the eigenvalue of h on ℓ by λ.*

(2) *There is a unique Jordan block of h associated to the eigenvalue λ. Denote its size by c.*

Let $R := \mathcal{O}_{\mathbb{P}(\mathcal{L})^h, x_0}$ be the local ring of the fixed point scheme $\mathbb{P}(\mathcal{L})^h$ at x_0. Then

$$R \cong k[X]/X^c.$$

Proof. Extend ℓ to a basis $\{\ell_0 = \ell, \ell_1, \dots, \ell_d\}$ of $\mathcal{L}$ such that the matrix $(h_{ij})_{0\le i,j\le d}$ of h under this basis is in the Jordan normal form. Under this basis, the point x_0 has projective coordinates $[X_0 : \cdots : X_d] = [1 : 0 : \cdots : 0] \in \mathbb{P}^d$. Let $Z_i = X_i/X_0$ $(1 \le i \le d)$ and let $\mathbb{A}^d$ be the affine space with coordinates $(Z_1, \dots, Z_d)$. Then we can identify the local ring of $\mathbb{P}^d$ at x_0 with the local ring of $\mathbb{A}^d$ at the origin. Since h fixes x_0, we know that h acts on the local ring of $\mathbb{A}^d$ at the origin (although h does not stabilize $\mathbb{A}^d$ in general). Since (h_{ij}) is in the Jordan normal form, we know that the action of h on the latter is given explicitly by

$$hZ_i = \frac{h_{i,i}X_i + h_{i,i+1}X_{i+1}}{h_{0,0}X_0 + h_{0,1}X_1} = \frac{h_{i,i}Z_i + h_{i,i+1}Z_{i+1}}{h_{0,0} + h_{0,1}Z_1}, \quad 1 \le i \le d,$$

where $h_{i,i+1}Z_{i+1}$ is understood as 0 when $i = d$. Hence the local equations at the origin of $\mathbb{A}^d$ which cut out the h-fixed point scheme are given by

$$(h_{0,0} - h_{i,i})Z_i + h_{0,1}Z_1Z_i = h_{i,i+1}Z_{i+1}, \quad 1 \le i \le d.$$

By hypothesis (2), we have $h_{0,0} - h_{i,i} \neq 0$ if and only if $i \ge c$. Thus when $i \ge c$, we know that $(h_{0,0} - h_{i,i}) + h_{0,1}Z_1$ is a unit in the local ring of $\mathbb{A}^d$ at the origin, and so Z_i can be solved as a multiple of $h_{i,i+1}Z_{i+1}$ when $i \ge c$. It follows that

$$Z_i = 0, \quad i \ge c.$$

If $c = 1$, then $Z_1 = \cdots = Z_d = 0$ and the local ring R in question is isomorphic to k as desired. If $c > 1$, then $h_{0,1} = 1$ and we find the equations for $i = 1, \cdots, c-1$ simplify to

$$Z_1Z_1 = Z_2,\ Z_1Z_2 = Z_3, \cdots, Z_1Z_{c-2} = Z_{c-1},\ Z_1Z_{c-1} = 0.$$

Hence the local ring R in question is isomorphic to (the localization at the ideal $(Z_1, Z_2, \cdots, Z_{c-1})$ or (Z_1) of)

$$k[Z_1, Z_2, \dots, Z_{c-1}]/(Z_1^2 - Z_2, Z_1^3 - Z_3, \cdots, Z_1^{c-1} - Z_{c-1}, Z_1^c) \cong k[Z_1]/Z_1^c,$$

as desired. □

Theorem 4.3.5. *Let $g \in J(\mathbb{Q}_p)$ be regular semisimple and minuscule. Let x_0 be a point in $(\delta(\mathcal{M}) \cap \mathcal{N}^g)(k)$. Also denote by x_0 the image of x_0 in $\mathcal{V}(\Lambda)(k)$ as in*

Proposition 4.1.2 and define λ and c as in Definition 4.3.1. Assume $p > c$. Then the complete local ring of $\delta(\mathcal{M}) \cap \mathcal{N}^g$ at x_0 is isomorphic to $k[X]/X^c$.

Proof. Let $\hat{\mathcal{S}}$ be the complete local ring of $\delta(\mathcal{M}) \cap \mathcal{N}^g$ at x_0. By Proposition 4.1.2 and by the fact that $\mathcal{V}(\Lambda)$ is smooth of dimension d (Section 2.5), we know that $\hat{\mathcal{S}}$ is a quotient of the power series ring $k[\![X_1, \cdots, X_d]\!]$. By Proposition 4.1.2, Lemma 4.2.5 and Proposition 4.3.3, we know that $\hat{\mathcal{S}}/\mathfrak{m}_{\hat{\mathcal{S}}}^p$ is isomorphic to $k[X]/X^c$ as a k-algebra. In such a situation, it follows from the next abstract lemma that $\hat{\mathcal{S}} \cong k[X]/X^c$. □

Lemma 4.3.6. *Let I be a proper ideal of $k[\![X_1, \cdots, X_d]\!]$ and let*

$$\hat{\mathcal{S}} = k[\![X_1, \cdots, X_d]\!]/I.$$

Let $\mathfrak{m}$ be the maximal ideal of $k[\![X_1, \cdots, X_d]\!]$ and let $\mathfrak{m}_{\hat{\mathcal{S}}}$ be the maximal ideal of $\hat{\mathcal{S}}$. Assume there is a k-algebra isomorphism $\beta : \hat{\mathcal{S}}/\mathfrak{m}_{\hat{\mathcal{S}}}^p \xrightarrow{\sim} k[X]/X^c$ for some integer $1 \le c < p$. Then $\hat{\mathcal{S}}$ is isomorphic to $k[X]/X^c$ as a k-algebra.

Proof. We first notice that if R_1 is any quotient ring of $k[\![X_1, \cdots, X_d]\!]$ with its maximal ideal $\mathfrak{m}_1$ satisfying $\mathfrak{m}_1 = \mathfrak{m}_1^2$ (i.e., R_1 has zero cotangent space), then $R_1 = k$. In fact, R_1 is noetherian and we have $\mathfrak{m}_1^l = \mathfrak{m}_1$ for all $l \in \mathbb{Z}_{\ge 1}$, so by Krull's intersection theorem we conclude that $\mathfrak{m}_1 = 0$ and $R_1 = k$.

Suppose $c = 1$. Then $\hat{\mathcal{S}}/\mathfrak{m}_{\hat{\mathcal{S}}}^p \cong k$, so $\hat{\mathcal{S}}$ has zero cotangent space and thus $\hat{\mathcal{S}} = k$ as desired. Next we treat the case $c \ge 2$. Let α be the composite

$$\alpha : k[\![X_1, \cdots, X_d]\!] \to \hat{\mathcal{S}}/\mathfrak{m}_{\hat{\mathcal{S}}}^p \xrightarrow{\beta} k[X]/X^c.$$

Let $J = \ker \alpha$. Since α is surjective, we reduce to prove that $I = J$. Note that because β is an isomorphism we have

$$I + \mathfrak{m}^p = J. \tag{4.3.6.1}$$

In the following we prove $\mathfrak{m}^p \subset I$, which will imply $I = J$ and hence the lemma. The argument is a variant of [RTZ, Lemma 11.1].

Let $Y \in k[\![X_1, \cdots, X_d]\!]$ be such that $\alpha(Y) = X$. Since X generates the maximal ideal in $k[X]/X^c$, we have

$$\mathfrak{m} = J + (Y). \tag{4.3.6.2}$$

Then by (4.3.6.1) and (4.3.6.2) we have $\mathfrak{m} = I + (Y) + \mathfrak{m}^p$, and so the local ring $k[\![X_1, \cdots, X_d]\!]/(I + (Y))$ has zero cotangent space. We have observed that the cotangent space being zero implies that the ring has to be k, or equivalently

$$\mathfrak{m} = I + (Y). \tag{4.3.6.3}$$

Now we start to show $\mathfrak{m}^p \subset I$. By (4.3.6.3) we have $\mathfrak{m}^p \subset I + (Y^p)$, so we only need to prove $Y^p \in I$. We will show the stronger statement that $Y^c \in I$. By Krull's intersection theorem, it suffices to show that $Y^c \in I + \mathfrak{m}^{pl}$ for all $l \geq 1$. In the following we show this by induction on l.

Assume $l = 1$. Note that $\alpha(Y^c) = 0$, so by (4.3.6.1) we have

$$Y^c \in J = I + \mathfrak{m}^p.$$

Suppose $Y^c \in I + \mathfrak{m}^{pl}$ for an integer $l \geq 1$. Write

$$Y^c = i + m, \quad i \in I,\ m \in \mathfrak{m}^{pl}. \tag{4.3.6.4}$$

By (4.3.6.2) we know

$$\mathfrak{m}^{pl} \subset (J + (Y))^{pl} \subset \sum_{s=0}^{pl} J^s (Y)^{pl-s}.$$

Thus we can decompose $m \in \mathfrak{m}^{pl}$ into a sum

$$m = \sum_{s=0}^{pl} j_s Y^{pl-s}, \quad j_s \in J^s. \tag{4.3.6.5}$$

By (4.3.6.4) and (4.3.6.5), we have

$$Y^c = i + \sum_{s=0}^{pl} j_s Y^{pl-s},$$

and so

$$Y^c - \sum_{s=0}^{pl-c} j_s Y^{pl-s} = i + \sum_{s=pl-c+1}^{pl} j_s Y^{pl-s}. \tag{4.3.6.6}$$

Denote

$$A := \sum_{s=0}^{pl-c} j_s Y^{pl-s-c}.$$

Then the left hand side of (4.3.6.6) is equal to $(1 - A)Y^c$. Hence we have

$$\begin{aligned}(1 - A)Y^c = i + \sum_{s=pl-c+1}^{pl} j_s Y^{pl-s} &\subset I + J^{pl-c+1} \\ &\xrightarrow{(4.3.6.1)} I + (I + \mathfrak{m}^p)^{pl-c+1} = I + \mathfrak{m}^{p(pl-c+1)} \subset I + \mathfrak{m}^{p(l+1)},\end{aligned}$$

where for the last inclusion we have used $c < p$. Since $1 - A$ is a unit in $k[\![X_1, \cdots, X_d]\!]$ (because $c < p$), we have $Y^c \in I + \mathfrak{m}^{p(l+1)}$. By induction, $Y^c \in I + \mathfrak{m}^{pl}$ for all $l \in \mathbb{Z}_{\geq 1}$, as desired. $\square$

Acknowledgments

We are very grateful to M. Rapoport and W. Zhang for helpful conversations and comments. Our debt to [Rapoport et al. 2013] should be clear to the readers.

References

[Berthelot and Ogus 1978] P. Berthelot and A. Ogus, *Notes on crystalline cohomology*, Princeton University Press, 1978. MR Zbl

[Gan et al. 2012] W. T. Gan, B. H. Gross, and D. Prasad, "Symplectic local root numbers, central critical L values, and restriction problems in the representation theory of classical groups", pp. 1–109 in *Sur les conjectures de Gross et Prasad, I*, Astérisque **346**, Société Mathématique de France, Paris, 2012. MR Zbl

[Kudla and Rapoport 2011] S. Kudla and M. Rapoport, "Special cycles on unitary Shimura varieties, I: Unramified local theory", *Invent. Math.* **184**:3 (2011), 629–682. MR Zbl

[Li and Zhu 2017] C. Li and Y. Zhu, "Arithmetic intersection on GSpin Rapoport–Zink spaces", preprint, 2017. arXiv

[Mihatsch 2016] A. Mihatsch, "Relative unitary RZ-spaces and the Arithmetic Fundamental Lemma", preprint, 2016. arXiv

[Rapoport and Zink 1996] M. Rapoport and T. Zink, *Period spaces for p-divisible groups*, Annals of Mathematics Studies **141**, Princeton University Press, 1996. MR Zbl

[Rapoport et al. 2013] M. Rapoport, U. Terstiege, and W. Zhang, "On the arithmetic fundamental lemma in the minuscule case", *Compos. Math.* **149**:10 (2013), 1631–1666. MR Zbl

[Vollaard 2010] I. Vollaard, "The supersingular locus of the Shimura variety for $\mathrm{GU}(1, s)$", *Canad. J. Math.* **62**:3 (2010), 668–720. MR Zbl

[Vollaard and Wedhorn 2011] I. Vollaard and T. Wedhorn, "The supersingular locus of the Shimura variety of $\mathrm{GU}(1, n-1)$, II", *Invent. Math.* **184**:3 (2011), 591–627. MR Zbl

[Zhang 2012] W. Zhang, "On arithmetic fundamental lemmas", *Invent. Math.* **188**:1 (2012), 197–252. MR Zbl

Communicated by Michael Rapoport
Received 2017-05-26 Revised 2017-09-06 Accepted 2017-10-05

lichaocs@gmail.com *Department of Mathematics, Columbia University, New York, NY, United States*

yihang@math.columbia.edu *Department of Mathematics, Columbia University, New York, NY, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 11 No. 10 2017

1937-0652(2017)11:10;1-7