EXCEPTIONAL SPLITTING OF REDUCTIONS OF ABELIAN SURFACES

ANANTH N. SHANKAR AND YUNQING TANG

Abstract. Heuristics based on the Sato–Tate conjecture and the Lang–Trotter philosophy suggest that an abelian surface defined over a number field has infinitely many places of split reduction. We prove this result for abelian surfaces with real multiplication. As in [Cha14] and [Elk89], this shows that a density-zero set of primes pertaining to the reduction of abelian varieties is infinite. The proof relies on the Arakelov intersection theory on Hilbert modular surfaces.

1. Introduction

1.1. **Infinitely many nonsimple reductions of a given abelian surface.** Murty and Patankar conjectured in [MP08] that an absolutely simple abelian variety over a number field has absolutely simple reduction for a density one set of primes (up to a finite extension) if and only if its endomorphism ring is commutative. Chavdarov ([Cha97]) proved their conjecture in the case of abelian varieties of dimension 2 or 6 whose geometric endomorphism ring is $\mathbb{Z}$ and Achter ([Ach12]) proved their conjecture in the case of abelian surfaces with real multiplication (i.e. when the endomorphism algebra of the surface contains a real quadratic field). Conditional upon the Mumford–Tate conjecture (which is known in the case of abelian surfaces), Zywina ([Zyw14]) established Murty and Patankar's conjecture in full generality.

It is natural to inquire whether the set of primes (conjecturally a density zero set!) at which a given abelian variety does not have absolutely simple reduction is finite or infinite. Based on the Sato–Tate conjecture and the Lang–Trotter philosophy for abelian surfaces (see §1.2), it is expected that the (density zero) set of places of nonsimple reduction of a simple abelian surface is infinite. The main result of this paper is the following:

Theorem 1. *Let A be an abelian surface over a number field K. Suppose that $F \subset \mathrm{End}(A) \otimes \mathbb{Q}$, where F is a real quadratic field. Then A modulo v is geometrically isogenous to the self-product of an elliptic curve*[1] *for infinitely many primes v of K.*

1.2. **A heuristic based on the Sato–Tate conjecture and the Lang–Trotter philosophy.** The classical Sato–Tate conjecture addresses the distribution of Frobenius elements involved in the Galois representation on the étale cohomology of a fixed elliptic curve defined over a number field. The work of Katz–Sarnak [KS99], Serre [Ser12], and Fité–Kedlaya–Rotger–Sutherland ([FKRS12]) generalizes this conjecture to higher dimensional abelian varieties. We focus on the case of abelian surfaces with real multiplication and offer a heuristic which indicates that such surfaces have infinitely many places of nonsimple reduction.

For simplicity, assume that A is an abelian surface defined over $\mathbb{Q}$ such that $\mathrm{End}(A)\otimes\mathbb{Q} = \mathrm{End}(A_{\overline{\mathbb{Q}}})\otimes\mathbb{Q} = \mathbb{Q}[\sqrt{D}] = F$, a real quadratic field. We roughly follow the idea of Lang–Trotter (see [LT76, Part I] for more refined computations and further details). For each prime ℓ of good reduction for A, the characteristic polynomial of the Frobenius endomorphism of A modulo ℓ is of the form $x^4 + a_1x^3 + a_2x^2 + \ell a_1 x + \ell^2$, with $a_1, a_2 \in \mathbb{Z}$. The roots of this polynomial come in complex conjugate pairs $\lambda_1, \overline{\lambda_1}, \lambda_2, \overline{\lambda_2}$, and each λ_i has absolute value $\ell^{1/2}$. Define $s_{i,\ell} = \frac{\lambda_i + \overline{\lambda_i}}{\sqrt{\ell}}$. The Sato–Tate conjecture for abelian surfaces with real multiplication predicts that the distribution of $(s_{1,\ell}, s_{2,\ell}) \in [-2,2]\times[-2,2]$, as ℓ varies, converges to the measure on $[-2,2]\times[-2,2]$ defined by the function $(\frac{1}{2\pi})^2\sqrt{4-s_1^2}\sqrt{4-s_2^2}$ (for instance, see [Ked15]). By Honda-Tate theory, A splits modulo ℓ if $s_{1,\ell} = s_{2,\ell}$.

An analogue of the probablistic model developed by Lang–Trotter in [LT76] can be used in our setting to estimate the probability that our abelian surface splits at ℓ. Since we are only aiming for a heuristic that suggests infinitude, from now on, our estimate is only up to an absolute constant. From the Sato–Tate

[1] Such elliptic curve depends on v.

measure for $(s_{1,\ell}, s_{2,\ell})$, we obtain a probability measure μ for $t_\ell = s_{1,\ell} - s_{2,\ell}$. Since $s_{1,\ell} - s_{2,\ell} \in \frac{\sqrt{D}}{\sqrt{\ell}}\mathbb{Z}$, we follow Lang–Trotter ([LT76, pp. 29–33]) to discretize the measure μ into a probability measure on the set $\frac{\sqrt{D}}{\sqrt{\ell}}\mathbb{Z} \cap [-4, 4]$. As in [LT76, pp. 31–32], the probability that $t_\ell = 0$ is $\asymp \frac{1}{\sqrt{\ell}}$. Since $\sum_{\ell \text{ prime}} \frac{1}{\sqrt{\ell}}$ diverges, there should be infinitely many primes ℓ such that $s_{1,\ell} = s_{2,\ell}$ and hence A should have infinitely many primes of nonsimple reduction.

Note that the above estimate is based on a probability model on isogeny classes of such abelian surfaces. One may also build a model based on estimating the number of isomorphism classes of simple and non-simple abelian surfaces, and obtain similar predictions. See for example the work of Achter and Howe [AH17, §2].

1.3. **Related results.** The Sato–Tate and Lang–Trotter conjectures for elliptic curves and pairs of elliptic curves also suggests that the set of primes in either of the following two situations is infinite:

(1) given an elliptic curve E over a number field, consider primes v such that $E \bmod v$ is supersingular;
(2) given a pair of non-isogenous elliptic curves E_1, E_2 over a number field, consider primes v such that the reductions of $E_1, E_2 \bmod v$ become geometrically isogenous.

As in our case, both sets of primes have density zero (after taking a finite extension). Indeed, Serre conjectured that up to taking a finite extension of the field of definition, a given abelian variety over a number field has ordinary reduction at a density-one set of primes. Katz proved Serre's conjecture in the case of elliptic curves and abelian surfaces ([Ogu82, pages 370–372]). Sawin (in [Saw16]) made explicit the smallest field extension that is required for abelian surfaces.

The second set also has density zero (after taking a finite extension) by an application of Faltings' isogeny theorem ([Fal86]) and a result of Serre on Frobenius tori (see, for instance, [Chi92] for the definitions): indeed, suppose that E_1 and E_2 are a pair of non-CM non-isogenous elliptic curves (if one or both of the curves have CM, a similar argument to the one given below would work). Then, by Faltings' isogeny theorem, the ℓ-adic algebraic monodromy group G_ℓ of $E_1 \times E_2$ contains $\mathrm{SL}_{2,\mathbb{Q}_\ell} \times \mathrm{SL}_{2,\mathbb{Q}_\ell}$. Serre's theory of Frobenius tori (see [Chi92, Corollary 3.8]) implies that the Frobenius element Frob_v generates a maximal torus of G_ℓ for a density one set of primes v (after taking a finite extension). By Honda–Tate theory, if E_1 is (geometrically) isogenous to E_2 modulo v, then the Frobenius eigenvalues of E_1 are the same as the Frobenius eigenvalues of E_2 (up to multiplication by roots of unity), which implies that the torus in G_ℓ generated by Frob_v cannot be a maximal torus.

Elkies proved (1) in [Elk87, Elk89] when the elliptic curve is defined over a number field with at least one real embedding and Charles proved (2) in [Cha14]. Theorem 1 is an analogue of these two results. Indeed, all three results establish that certain thin sets of primes related to the reduction of abelian varieties are infinite.

1.4. **The strategy of the proof.** The proof of Theorem 1 builds on the idea of the proof of the main theorem in [Cha14], where Charles uses Arakelov intersection theory on the modular curve $X_0(1)$ to prove his result. In our case, we use Arakelov intersection theory on the Hilbert modular surface $\mathcal{H}$ (see §2.1 for the precise definition).

Let $[A] \in \mathcal{H}(K)$ denote the point determined by A. Loosely speaking, the modular curve embeds canonically into $\mathcal{H}$ (we label its image Δ) and parametrizes the locus of split abelian surfaces (along with the product polarization). A natural strategy is to consider the Arakelov intersection of Δ with Hecke orbits of $[A]$.[2] The local contribution at a finite prime v is positive precisely when the Hecke orbit of $[A]$ intersects Δ modulo v. The reduction of A modulo v would be geometrically nonsimple for such v.

In our proof, we replace Δ with a compact Hirzebruch–Zagier divisor $\mathcal{T}$ (see §2.1 and §5.1 for definitions). Hirzebruch–Zagier divisors of $\mathcal{H}$ have the feature that the abelian surface B admits quaternionic multiplication if $[B]$ lies on these divisors. This has the consequence that over a finite field, an abelian surface is not geometrically simple if it lies on a special divisor (see Corollary 2.1.7). There are two advantages of using a *compact* Hirzebruch–Zagier divisor $\mathcal{T}$: the first is that we do not have to deal with places of bad reduction for A. The second is that we are able to avoid all the cusps of $\mathcal{H}$ in the archimedean contribution to the global Arakelov intersection.

[2]Here we refer to the 1-cycle, given by taking the Zariski closure of the Hecke orbit of $[A]$ in $\mathcal{H}$ over $\operatorname{Spec}\mathbb{Z}$.

In order to prove Theorem 1, it would suffice to prove that the set of primes which contribute to the intersection is infinite as we vary over infinitely many well-chosen Hecke orbits of $[A]$. There are two steps involved in proving this:

- A local step, where we bound the local contribution of the intersection at every place.
- A global step, where we compute the growth of the Arakelov intersection of Hecke orbits of $[A]$ with $\mathcal{T}$. The growth is expressed in terms of the degree of the Hecke operators, and is seen to grow asymptotically faster than the local contributions at each place.

Consequently, it follows that more and more primes contribute to this intersection as we vary the Hecke orbit of A.

The abelian surfaces parametrized by Hirzebruch–Zagier divisors have extra special endomorphisms (recalled in §2.1). In the non-archimedean case, our methods are very different from the ones used in [Cha14]. For a finite place v, we use Grothendieck–Messing theory to prove statements about the rate of decay of special endomorphisms of $A[\ell^\infty]$ modulo higher and higher powers of v. This method avoids the use of CM lifts and can be used in other Arakelov-theoretic situations. We use these results and Geometry-of-numbers arguments to bound the number of special endomorphisms of A modulo powers of v. This allows us to prove that the v-adic contribution grows asymptotically slower than the global intersection for most of the Hecke orbits that we consider. Indeed, if there were too many Hecke orbits $T_{\mathfrak{p}}([A])$ having large v-adic intersection with $\mathcal{T}$, then A modulo v^n would have too many special endomorphisms as $n \to \infty$.

The arguments used to bound the archimedean contribution are very different from the ones used to bound the finite contributions. A key step in bounding the archimedean contribution is the following statement: for a fixed infinite place, if $[A]$ is close to two Hirzebruch–Zagier divisors, then $[A]$ must be close to their intersection, which is a CM abelian surface.

In order to prove the global part of our result, it is necessary to relate the global Arakelov intersection of $\mathcal{T}$ and certain Hecke orbits $T_{\mathfrak{p}}([A])$ (see §2.2.1 for the precise definition) to the intersection of $[A]$ and $\mathcal{T}$. We accomplish this in two steps:

- We use Borcherds' theory (briefly recalled after Lemma 5.1.1) to construct a compact special divisor, whose class in the Picard group of a toroidal compactification of $\mathcal{H}$ equals, up to multiplying by a constant in $\mathbb{Z}_{>0}$, the class of the Hodge bundle. Consequently, the global Arakelov intersection of $[A]$ with $\mathcal{T}$ (endowed with a suitable Hermitian metric),[3] up to multiplying by a suitable constant, equals the Faltings height of A.
- We relate the Faltings height of $T_{\mathfrak{p}}([A])$ to the Faltings height of A when A has potentially good reduction at p in Proposition 5.1.6. This extends a result of Autissier ([Aut05, Theorem 5.1]), which only applies to A with potentially ordinary reduction at p.

It follows that the global intersection number grows faster than any local contribution. Hence, infinitely many primes occur in the intersection of $T_{\mathfrak{p}}([A])$ and $\mathcal{T}$ as $p \to \infty$.

1.5. **Organization of the paper.** In §2, we recall the definitions of the Hilbert modular surface and the Hirzebruch–Zagier divisors. In §3, we bound the archimedean contribution. We spend §4 counting special endomorphisms and bounding the non-archimedean contribution. We use Borcherds' theory in §5.1 to choose a compact Hirzebruch–Zagier divisor and relate the Arakelov intersection number to Faltings height. We also extend Autissier's result to the setting of Hilbert modular surfaces. Finally, we assemble all these results together in §5.2 to prove Theorem 1.

1.6. **Notation and conventions.** We use K to denote a number field and let $\mathcal{O}_K$ be its ring of integers. As in Theorem 1, we use F to denote a fixed real quadratic field with discriminant D; its ring of integers is denoted by $\mathcal{O}_F$ and $\mathfrak{d}_F$ is its different ideal. For $a \in F$, we use $\operatorname{Nm} a$ to denote its $F/\mathbb{Q}$-norm.

The statement of Theorem 1 is invariant under isogeny. Therefore, we will always assume that A has real multiplication by the maximal order $\mathcal{O}_F$, and is equipped with an $\mathfrak{a}$-polarization for some (integral) ideal $\mathfrak{a}$ of $\mathcal{O}_F$; see [Pap95, § 2.1 item 2 before Def. 2.1.1] for the definition of an $\mathfrak{a}$-polarization. We may also assume that A has semistable reduction over K. For any abelian varieties B, B', we use $\operatorname{End}(B)$ and $\operatorname{Hom}(B, B')$ to denote the endomorphism ring of B and the $\mathbb{Z}$-module given by homomorphisms from B to B'.

[3] Technically speaking, since we will not make $[A]$ into an arithmetic cycle, here by Arakelov intersection, we mean the height of the 1-cycle $[A]$ with respect to the Arakelov divisor $\mathcal{T}$, which is endowed with a Hermitian metric by Borcherds' theory.

Let $\mathcal{H}$ be the Hilbert modular surface over $\mathbb{Z}$ associated to F which is the moduli stack of abelian surfaces B with real multiplication by $\mathcal{O}_F$ and an $\mathfrak{a}$-polarization. This is a Deligne–Mumford stack and we use $[B]$ to denote the point of $\mathcal{H}$ determined by B. Sometimes, we may denote a point of $\mathcal{H}$ by $[B']$; this means that B' is the abelian surface that determines this given point.

We always use mathcal letters to mean the natural extension over certain ring of integers. For example, we use $\mathcal{A}$ to denote the everywhere semistable semi-abelian scheme over $\mathcal{O}_K$ such that $\mathcal{A}_K = A$.

Throughout the text, v means a place of K, either archimedean or finite. If v is finite, we use $\mathbb{F}_v$ to denote its residue field and e_v to denote the degree of ramification of K at v. If $\mathcal{A}$ has good reduction at v, we use $\mathcal{A}_{v,n}$ to denote $\mathcal{A}$ modulo v^n. We always use p to denote a prime number which is totally split in the narrow Hilbert class field of F and denote by $\mathfrak{p}, \mathfrak{p}'$ the two primes ideals of F above p. We use $A[p]$ and $A[\mathfrak{p}]$ to denote the p-torsion and $\mathfrak{p}$-torsion subgroups of A.

Acknowledgements. We would like to thank Jeff Achter, George Boxer, Francesc Castella, Kęstutis Česnavičius, François Charles, William Chen, Noam Elkies, Ziyang Gao, Chi-Yun Hsu, Nicholas Katz, Ilya Khayutin, Djordjo Milovic, Lucia Mocz, Peter Sarnak, William Sawin, Arul Shankar, Jacob Tsimerman, Tonghai Yang, and Shou-Wu Zhang for useful comments and/or discussions. We are also grateful to Kęstutis Česnavičius, Chao Li, Mark Kisin, Lucia Mocz, and Salim Tayou for very useful comments on previous versions of this paper. We would like to thank Davesh Maulik for pointing out a gap in Theorem 4.1.1 in an earlier version of this paper. Y. T., during her stay at the Institute for Advanced Study, was supported by the NSF grant DMS-1128115 to IAS. Finally, it is a pleasure to thank the anonymous referee for detailed comments and suggestions, and also for helping to improve both the mathematical content as well as the exposition in this paper.

2. Hirzebruch–Zagier divisors and Hecke orbits

In this section, we first recall the definition of Hirzebruch–Zagier divisors and their properties and then we specify the Hecke orbits that will be used in the rest of the paper.

2.1. **The Hilbert modular surface and the Hirzebruch–Zagier divisors.** Recall that we use $\mathcal{H}$ to denote the moduli stack over $\operatorname{Spec}\mathbb{Z}$ that parametrizes abelian surfaces with real multiplication by $\mathcal{O}_F$ and an $\mathfrak{a}$-polarization (see [Pap95, Def. 2.1.1]). It is a Deligne–Mumford stack. A totally positive element $a \in \mathfrak{a}$ gives rise to a polarization on A and a symplectic form ψ on the Betti cohomology group $W = H^1_B(A(\mathbb{C}), \mathbb{Q})$ (here we choose an embedding $K \to \mathbb{C}$). The set of $\mathrm{GSp}(W,\psi)(\mathbb{R})$-conjugate cocharacters of the Hodge cocharacter (from the Hodge decomposition of $W \otimes \mathbb{C}$) of $A_{\mathbb{C}}$ coincides with $\mathbb{H}_2^{\pm}$, the upper and lower Siegel half plane of genus 2. Let $G \subset \mathrm{GSp}(W,\psi)$ be the subgroup (over $\mathbb{Q}$) that commutes with $\mathcal{O}_F \subset \mathrm{End}(W)$ and let $G_1 = \mathrm{Res}^F_{\mathbb{Q}} \mathrm{GL}_2$. Then G is naturally isomorphic to the following subgroup of G_1: for any $\mathbb{Q}$-algebra R, the set $G(R)$ consists of matrices with determinant in R (instead of $R \otimes_{\mathbb{Q}} F$). The embedding $G \subset \mathrm{GSp}_4$ induces an embedding[4] of the Hilbert modular surface $\mathcal{H}_{\mathbb{Q}} = Sh(G, X)$ into $Sh(\mathrm{GSp}(W,\psi), \mathbb{H}_2^{\pm})$ (see [vdG88, Chp. IX.1].) Here X is the subset of $\mathbb{H}_2^{\pm}$ which consists of cocharacters conjugate to the Hodge cocharacter of $A_{\mathbb{C}}$ under $G(\mathbb{R})$.

Let $\overline{\mathcal{H}}^{\mathrm{tor}}$ be a toroidal compactification of $\mathcal{H}$ as in [Rap78] (see also [Cha90, §3]). The stack $\overline{\mathcal{H}}^{\mathrm{tor}}$ is regular and proper and $\overline{\mathcal{H}}^{\mathrm{tor}} \backslash \mathcal{H}$ is a normal crossing divisor (see, for example, [Pap95, 2.1.2, 2.1.3] for the regularity of $\mathcal{H}$ and [Cha90, Thm. 3.6, 4.3] for the statements about the boundary). Hence the arithmetic intersection theory developed by Burgos Gil, Kramer, and Kühn in [BGKK07] applies to $\overline{\mathcal{H}}^{\mathrm{tor}}$ (see, for example, [BBGK07, §1, §6] for a summary of their theory). We will use $[\mathcal{A}]$ (resp. $[A]$) to denote the unique $\mathcal{O}_K$-point (resp. K-point) of $\overline{\mathcal{H}}^{\mathrm{tor}}$ corresponding to A (the stack $\overline{\mathcal{H}}^{\mathrm{tor}}$ being proper allows us to do this).[5]

2.1.1. We now summarize some basic facts about $\mathcal{H}_{\mathbb{C}}$ and the Hirzebruch–Zagier divisors. The facts discussed here can be found in [vdG88, Chp. I, V, IX], [BBGK07, §2.3, §5.1] and [Gor02, Chp. 2]; however, since conventions differ, we will use this subsection to fix our notation. After defining Hirzebruch–Zagier divisors,

[4] One needs to choose suitable level structure to ensure that the natural finite morphism is an embedding.

[5] In general, one needs to pass to a finite field extension to extend a K-point on a proper Deligne–Mumford stack to an $\mathcal{O}_K$ point; however, since we have assumed that A has a semistable integral model $\mathcal{A}$ over $\mathcal{O}_K$, we do not need to pass to a further field extension.

we first show that among them there exists a nonempty compact one, and then we give a moduli interpretation of these divisors.

Let $\mathbb{H}$ denote the upper half plane. The two real embeddings of F induce two embeddings $\sigma_1, \sigma_2 : \mathrm{SL}_2(F) \to \mathrm{SL}_2(\mathbb{R})$. The action of $g \in \mathrm{SL}_2(F)$ on $\mathbb{H}^2$ is given by $\sigma_1(g)$ on the first copy of $\mathbb{H}$ and by $\sigma_2(g)$ on the second copy. We have $\mathcal{H}(\mathbb{C}) = \Gamma\backslash\mathbb{H}^2$, where $\Gamma = \mathrm{SL}_2(F) \cap \begin{pmatrix} \mathcal{O}_F & (\mathfrak{a}\mathfrak{d}_F)^{-1} \\ \mathfrak{a}\mathfrak{d}_F & \mathcal{O}_F \end{pmatrix}$.[6]

For any $r \in \mathbb{Z}_{>0}$, we recall the definition of the Hirzebruch–Zagier divisors $T(r)$ in $\mathcal{H}_{\mathbb{C}}$ (see for example [vdG88, V.1.3] and [BBGK07, sec. 2.3]). Let γ' denote the $\mathrm{Gal}(F/\mathbb{Q})$-conjugate of a given $\gamma \in F$. Consider the lattice

$$L = \left\{ \begin{pmatrix} a & \gamma \\ \gamma' & b \end{pmatrix} : a \in (D \operatorname{Nm} \mathfrak{a})\mathbb{Z}, b \in \mathbb{Z}, \gamma \in \mathfrak{a} \right\}$$

in the rational quadratic space

$$V = \left\{ \begin{pmatrix} a & \gamma \\ \gamma' & b \end{pmatrix} : a, b \in \mathbb{Q}, \gamma \in F \right\}$$

with the quadratic form given by the determinant. The group Γ acts on V via $v.g = (g')^t \cdot v \cdot g$ for $g \in \Gamma, v \in V$ and this action preserves L. The quadratic space V is of signature $(2,2)$. One may also view $\mathcal{H}$ as an orthogonal type Shimura variety defined by $\mathrm{SO}(V)$. The divisor $T(r)$ is defined to be the reduced divisor in $\mathcal{H}_{\mathbb{C}}$ whose set of $\mathbb{C}$-points is the image of[7]

$$\bigcup_{M \in L,\, \det(M) = r \operatorname{Nm} \mathfrak{a}} \{(z_1, z_2) \in \mathbb{H}^2 : a z_1 z_2 + \gamma z_1 + \gamma' z_2 + b = 0\}.$$

Proposition 2.1.2. *The divisor $T(r)$ is nonempty if and only if $r \operatorname{Nm} \mathfrak{a}$ modulo $D \operatorname{Nm} \mathfrak{a}$ is $-\operatorname{Nm} \gamma$ for some $\gamma \in \mathfrak{a}$. In this case, $T(r)$ is defined over $\overline{\mathbb{Q}}$ and is either a modular curve or a Shimura curve defined by the indefinite quaternion algebra $\left(\frac{D, -r \operatorname{Nm} \mathfrak{a}}{\mathbb{Q}}\right)$. If r is not the norm of an ideal of $\mathcal{O}_F$, then $T(r)$ is a Shimura curve, and hence compact.*

Proof. The first assertion follows from the definition of a Hirzebruch–Zagier divisor. By the discussion on [vdG88, pp. 89–90], the divisor $T(r)$ is the union of Shimura curves defined by the quaternion algebra mentioned above and hence is defined over $\overline{\mathbb{Q}}$. The last assertion follows from [vdG88, Chp. V, 1.7]. □

Corollary 2.1.3. *Let q denote a rational prime inert in F. Then the divisor $T(qD)$ is non-empty and compact.*

Proof. As q is inert, qD is not the norm of an ideal of $\mathcal{O}_F$. Further, $qD \operatorname{Nm} \mathfrak{a}$ is 0 modulo $D \operatorname{Nm} \mathfrak{a}$. It follows from Proposition 2.1.2 that $T(qD)$ is compact and nonempty. □

Hirzebruch–Zagier divisors parametrize abelian surfaces with extra special endomorphisms. After recalling the definition of special endomorphisms, we sketch a proof of this fact (see Lemma 2.1.6),[8] which may be well known to experts. From now on, B denotes an abelian surface over some $\mathbb{Z}$-algebra with an $\mathfrak{a}$-polarization and $\mathcal{O}_F \subset \mathrm{End}(B)$. We fix a totally positive element in $\mathfrak{a} \cap \mathbb{Q}$ which provides a fixed polarization on B. We use this polarization to define the Rosati involution $(-)^*$ on $\mathrm{End}(B) \otimes \mathbb{Q}$.

Definition 2.1.4 (see also [KR99, Def. 1.2])**.** An element $s \in \mathrm{End}(B)$ is a *special endomorphism* if $a \circ s = s \circ a'$ for all $a \in \mathcal{O}_F \subset \mathrm{End}(B)$ and $s^* = s$.

The special endomorphisms of B form a sub-$\mathbb{Z}$-module of $\mathrm{End}(B)$. It is well known that the rank of this submodule is at most 4 (see for instance [HY12, Cor. 3.1.4]). The following lemma recalls the discussion after [KR99, Def. 1.2].

---

[6] The Hilbert modular surface $\mathcal{H}$ is connected and hence we may use $\mathrm{Res}^F_{\mathbb{Q}} \mathrm{SL}_2$ instead of G to study the complex points. Using the notation in [BBGK07], we work with $\Gamma(\mathcal{O}_F \oplus \mathfrak{a}\mathfrak{d}_F)$.

[7] This is the definition in [BBGK07]. The lattice in [vdG88] differs by a multiple of the scalar matrix $\sqrt{D} \cdot I$, so these two definitions of $T(r)$ coincide.

[8] We only deal with $T(Dr)$ since these are the divisors that we will use in the proof of Theorem 1. However, after minor modifications, the proof shows that any $T(r)$ parametrizes abelian surfaces with an extra special endomorphism.

Lemma 2.1.5. *Every special endomorphism s has the property that $s \circ s \in \mathbb{Z} \cdot \mathrm{Id}_B \subset \mathrm{End}(B)$, and hence $\mathrm{Deg}(s) = (s \circ s)^2$. The integer valued function $Q(s) = s \circ s$ is a positive definite quadratic form on the $\mathbb{Z}$-module of special endomorphisms of B.*

The $\mathbb{Q}$-vector space in $\mathrm{End}(B) \otimes \mathbb{Q}$ generated by special endomorphisms depends on the choice of the polarization on B. However, there are natural isomorphisms between the $\mathbb{Q}$-vector spaces of special endomorphisms defined by different polarizations and the quadratic forms coincide up to multiplying by a fixed scalar determined by the polarizations.

Lemma 2.1.6. *The Hirzebruch–Zagier divisor $T(Dr)$ defined in 2.1.1 is the locus of $\mathcal{H}_{\overline{\mathbb{Q}}}$ where the abelian surface has a special endomorphism s with $Q(s) = r \operatorname{Nm} \mathfrak{a}$. In particular, the degree of the endomorphism s is $(r \operatorname{Nm} \mathfrak{a})^2$.*

Proof. We only need to check the statement over $\mathbb{C}$. Given a point in $\mathcal{H}(\mathbb{C})$ corresponding to $(z_1, z_2) \in \mathbb{H}^2$, it corresponds to an abelian surface B with $B(\mathbb{C}) = \mathbb{C}^2/(\mathcal{O}_F(z_1, z_2) + (\mathfrak{a}\mathfrak{d}_F)^{-1})$. Any endomorphism of B is given by the induced map on $B(\mathbb{C})$ of some $\mathbb{C}$-linear map on $\mathbb{C}^2$. For any endomorphism s, the condition $f \circ s = s \circ f'$ for all $f \in \mathcal{O}_F$ is equivalent to the condition that the $\mathbb{C}$-linear map corresponding to s is of the form $(1,0) \mapsto (0, \alpha' z_2 + \beta'), (0,1) \mapsto (\alpha z_1 + \beta, 0)$ where $\alpha \in \mathfrak{a}\mathfrak{d}_F, \beta \in \mathcal{O}_F$. This linear map gives rise to an endomorphism of B if and only if the image of (z_1, z_2) is in the period lattice. In other words, there exists $\nu \in \mathcal{O}_F, \delta \in (\mathfrak{a}\mathfrak{d}_F)^{-1}$ such that $(z_2(\alpha z_1 + \beta), z_1(\alpha' z_2 + \beta')) = (\nu z_1 + \delta, \nu' z_2 + \delta')$.

For every component of $T(Dr)$, there exists $M \in L$ in 2.1.1 satisfying $\det(M) = Dr \operatorname{Nm} \mathfrak{a}$. Write $M = \begin{pmatrix} a & \gamma \\ \gamma' & b \end{pmatrix}$ where $a \in (D \operatorname{Nm} \mathfrak{a})\mathbb{Z}, b \in \mathbb{Z}, \gamma \in \mathfrak{a}$. Since $D \operatorname{Nm} \mathfrak{a} \mid \gamma\gamma'$, we have $\frac{\gamma}{\sqrt{D}} \in \mathfrak{a} \subset \mathcal{O}_F$. Moreover $\frac{a}{\sqrt{D}} \in (\operatorname{Nm} \mathfrak{a})\mathfrak{d}_F \subset \mathfrak{a}\mathfrak{d}_F$. We take $\alpha = \frac{a}{\sqrt{D}}$ and $\beta = \frac{\gamma'}{\sqrt{D}}$. Given $(z_1, z_2) \in \mathbb{H}^2$ such that $a z_1 z_2 + \gamma z_1 + \gamma' z_2 + b = 0$, we have

$$\alpha z_1 z_2 + \beta z_2 = \beta' z_1 - \frac{b}{\sqrt{D}}, \quad \alpha' z_1 z_2 + \beta' z_1 = \beta z_2 + \frac{b}{\sqrt{D}}.$$

Hence $(1,0) \mapsto (0, \alpha' z_2 + \beta'), (0,1) \mapsto (\alpha z_1 + \beta, 0)$ is an endomorphism s with $f \circ s = s \circ f'$ for all $f \in \mathcal{O}_F$.

To check that $s = s^*$, it is equivalent to check that for any $u, v \in H_1(B, \mathbb{Z})$, one has $E(su, v) = E(u, sv)$ where E is the Riemann form on $H^1(B, \mathbb{Z})$. Up to multiplying by a constant $\in \mathbb{Q}_{>0}$, E is the pull back of the standard alternating form ($\mathrm{Tr}_{F/\mathbb{Q}}$ of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$) on $\mathcal{O}_F \oplus (\mathfrak{a}\mathfrak{d}_F)^{-1}$ via the isomorphism $\mathcal{O}_F(z_1, z_2) + (\mathfrak{a}\mathfrak{d}_F)^{-1} \cong \mathcal{O}_F \oplus (\mathfrak{a}\mathfrak{d}_F)^{-1}$; see, for example, [vdG88, p. 208] and [BBGK07, the discussion after Thm. 5.1]. Since we have already checked that $f \circ s = s \circ f'$ for all $f \in \mathcal{O}_F$, one only needs to check the above equality for $u, v \in \{e_1 = (z_1, z_2), e_2 = (1,1)\} \subset \mathbb{C}^2$. By construction, $s e_1 = \beta' e_1 - \frac{b}{\sqrt{D}} e_2$ and $s e_2 = \alpha e_1 + \beta e_2$. The facts that $\mathrm{Tr}_{F/\mathbb{Q}} \beta = \mathrm{Tr}_{F/\mathbb{Q}} \beta'$, $\mathrm{Tr}_{F/\mathbb{Q}}(-\frac{b}{\sqrt{D}}) = 0$, and $\mathrm{Tr}_{F/\mathbb{Q}} \alpha = 0$ allow us to conclude that $s = s^*$. Moreover, on $\mathbb{C}^2$, the composite $s \circ s = \frac{\det(M)}{D} \cdot \mathrm{Id}_{\mathbb{C}^2}$. Hence s is a special endomorphism with $Q(s) = r \operatorname{Nm} \mathfrak{a}$.

On the other hand, the moduli space of B with a special endomorphism is 1-dimensional. Hence the two conditions $z_2(\alpha z_1 + \beta) = \nu z_1 + \delta$ and $z_1(\alpha' z_2 + \beta') = \nu' z_2 + \delta'$ are linearly dependent. Hence either $\alpha, \delta \in \mathbb{Q}$, $\beta = -\nu'$ or $\alpha \cdot \sqrt{D}$, $\delta \cdot \sqrt{D} \in \mathbb{Q}$, $\beta = \nu'$. In both cases, note that the special endomorphism s given by $(1,0) \mapsto (0, \alpha' z_2 + \beta'), (0,1) \mapsto (\alpha z_1 + \beta, 0)$ has the property that $s \circ s = (\alpha'\delta + \beta\beta') \cdot \mathrm{Id}_B$. In the first case, we have

$$\alpha z_1 z_2 + \beta' z_1 + \beta z_2 - \delta = 0, \beta\beta' + \alpha\delta = r \operatorname{Nm} \mathfrak{a} > 0,$$

and so there is no (z_1, z_2) satisfying the above condition (see for example [vdG88, V.4]). In the second case, take $a = \alpha \cdot \sqrt{D}, b = -\delta \cdot \sqrt{D}, \gamma = \beta' \cdot \sqrt{D}$. Then $M = \begin{pmatrix} a & \gamma \\ \gamma' & b \end{pmatrix} \in L$ and hence $[B] \in T(Dr)$. □

Let $\mathcal{T}(r)$ be the Zariski closure of $T(r)$ in $\overline{\mathcal{H}}^{\mathrm{tor}}$ over $\operatorname{Spec} \mathbb{Z}$.

Corollary 2.1.7. *Assume $T(Dr)$ is compact. Then for any finite place v, the points on $\mathcal{T}(Dr)_{\overline{\mathbb{F}}_v}$ correspond to abelian surfaces which are not absolutely simple. The abelian surfaces parametrized by $\mathcal{T}(Dr)$ admit a special endomorphism s such that $Q(s) = r \operatorname{Nm} \mathfrak{a}$.*

Proof. Since $T(Dr)$ is compact, every point parametrized by it has potentially good reduction. For any given point parametrized by $\mathcal{T}(Dr)_{\mathbb{F}_v}$, let $[\mathcal{B}]$ be an $\mathcal{O}_K$-valued point of $\mathcal{T}(Dr)$ lifting it, where K is some number

field. By Lemma 2.1.6, the endomorphism algebra of $\mathcal{B}_K$, and therefore of $\mathcal{B}_{\mathbb{F}_v}$, contains a quaternionic algebra which is split at infinity. This implies that $\mathcal{B}_{\mathbb{F}_v}$ is either supersingular or ordinary. In the supersingular case, $\mathcal{B}_{\mathbb{F}_v}$ is isogenous to E^2 for supersingular E, in which case we are done. On the other hand, the endomorphism algebra of a geometrically simple ordinary abelian variety has to be commutative. Indeed, the endomorphism algebra of an ordinary abelian variety is the same as that of its canonical lift (and thus, the canonical lift has CM), and every geometrically simple CM abelian variety in characteristic zero has commutative endomorphism ring.

The last assertion follows from Lemma 2.1.6 and the fact that the canonical reduction map $\mathrm{End}(\mathcal{B}_K) \to \mathrm{End}(\mathcal{B}_{\mathbb{F}_v})$ is injective. □

2.2. **Hecke orbits.** The idea of the proof of Theorem 1 is to show that the corresponding Hecke orbits of $[\mathcal{A}]$ intersect certain Hirzebruch–Zagier divisors at more and more places of K as one varies over certain well-chosen Hecke operators.[9] In this subsection, we specify the Hecke orbits which we will use later.

2.2.1. Recall that p is a prime which splits completely in the narrow Hilbert class field of F and $(p) = \mathfrak{p}\mathfrak{p}' \subset \mathcal{O}_F$. Hence $\mathfrak{p} = (\lambda)$, $\mathfrak{p}' = (\lambda')$ with $\lambda, \lambda' \in F$ totally positive and $\lambda\lambda' = p$. Let G_1^{ad} be the adjoint group of $G_1 = \mathrm{Res}^F_{\mathbb{Q}} \mathrm{GL}_2$. We denote by $G_1^{\mathrm{ad}}(\mathbb{R})_1$ the image of $G_1(\mathbb{R})$ in $G_1^{\mathrm{ad}}(\mathbb{R})$, and let $G_1^{\mathrm{ad}}(\mathbb{Q})_1$ be $G_1^{\mathrm{ad}}(\mathbb{Q}) \cap G_1^{\mathrm{ad}}(\mathbb{R})_1$. Since λ is totally positive, the image of the diagonal matrix $g_{\mathfrak{p}} := \mathrm{diag}(1, \lambda)$ under $G_1 \to G_1^{\mathrm{ad}}$ lies in $G_1^{\mathrm{ad}}(\mathbb{Q})_1$, so it induces a correspondence $T_{\mathfrak{p}}$ on $\mathcal{H}_{\mathbb{Z}[1/p]}$ (defined in [Del79]; see also [Kis10, sec. 3.2]).[10] The following lemma provides a moduli interpretation of $T_{\mathfrak{p}}$.

Lemma 2.2.2. *We have $\#T_{\mathfrak{p}}[A] = p + 1$. Over $\mathbb{Z}[1/p]$, the set $T_{\mathfrak{p}}[\mathcal{A}]$ consists of those points on $\mathcal{H}$ that correspond to a quotient of A by an order p subgroup in $A[\mathfrak{p}]$ endowed with the induced $\mathcal{O}_F$-structure and a suitable $\mathfrak{a}$-polarization.*[11]

Proof. The first assertion follows from the definitions: $\#T_{\mathfrak{p}}[A] = \#\Gamma/(g_{\mathfrak{p}}^{-1}\Gamma g_{\mathfrak{p}} \cap \Gamma) = \#\mathbb{P}^1(\mathbb{F}_p) = p + 1$.

For the second assertion, since the correspondence $T_{\mathfrak{p}}$ is étale, we only need to show the same statement for $T_{\mathfrak{p}}[A]$ over $\mathbb{C}$ for a fixed embedding of K into $\mathbb{C}$. On the one hand, as $\mathcal{O}_F$ acts on $A[\mathfrak{p}]$ via $\mathcal{O}_F/\mathfrak{p} \cong \mathbb{F}_p$ and $\mathbb{Z} \subset \mathcal{O}_F$ surjects onto $\mathbb{F}_p$, any subgroup of $A[\mathfrak{p}]$ is $\mathcal{O}_F$-invariant. Therefore, any quotient of A by an order p subgroup of $A[\mathfrak{p}]$ has the induced $\mathcal{O}_F$-structure. Moreover, by [BBGK07, Lem. 5.9], any such quotient of A is $\mathfrak{a}$-polarizable.

On the other hand, by [vdG88, p. 208], if a point $(z_1, z_2) \in \mathbb{H}^2$ corresponds to $A_{\mathbb{C}}$, then $A(\mathbb{C})$ is isomorphic to $\mathbb{C}^2/(\mathcal{O}_F(z_1, z_2) + (\mathfrak{a}\mathfrak{d}_F)^{-1})$. So $\mathrm{diag}(1, \lambda)z$ corresponds to $\mathbb{C}^2/(\mathcal{O}_F(z_1/\lambda, z_2/\lambda') + (\mathfrak{a}\mathfrak{d}_F)^{-1})$, hence the kernel of the isogeny defined by $\mathrm{diag}(1, \lambda)$ is contained in $\ker(\lambda) = A[\mathfrak{p}]$. On the quotient $\mathbb{C}^2/(\mathcal{O}_F(z_1/\lambda, z_2/\lambda') + (\mathfrak{a}\mathfrak{d}_F)^{-1})$, the $\mathcal{O}_F$-structure is the induced one and the choice of λ determines the $\mathfrak{a}$-polarization. Since the other elements in $T_{\mathfrak{p}}$ differ from $\mathrm{diag}(1, \lambda)$ by the action of some element in Γ (on $\mathbb{H}^2$ and on $A[\mathfrak{p}]$), the set $T_{\mathfrak{p}}$ injects into the set of order p subgroups of $A[\mathfrak{p}]$. Since both sets have cardinality $p + 1$, this is in fact a bijection. □

3. Archimedean places and equidistribution of Hecke orbits

Let Ψ be a meromorphic Hilbert modular form of parallel weight k over $\overline{\mathbb{Q}}$ such that $\mathrm{Div}(\Psi)$ in $\overline{\mathcal{H}}^{\mathrm{tor}}_{\overline{\mathbb{Q}}}$ is given by $\sum_{r \in \mathbb{I}} c_r T(r)$, where $k \in \mathbb{N}_{>0}$, $\mathbb{I}$ is a finite set, $c_r \in \mathbb{Z}$, $T(r)$ is compact and $D|r$ for all $r \in \mathbb{I}$. We further assume that $r = Dq$, where q is an inert prime in F. In the proof of Theorem 1, we will use Lemma 5.1.1 to construct such a meromorphic Hilbert modular form. Note that Ψ does not have poles or zeros along the boundary of $\overline{\mathcal{H}}^{\mathrm{tor}}_{\overline{\mathbb{Q}}}$.

We assume that $\mathrm{End}(A_{\overline{K}}) = \mathcal{O}_F$ and hence $T_{\mathfrak{p}}([A])$ does not intersect $T(r)$ in characteristic zero. This is the key case in the proof of Theorem 1. Fix an embedding $\sigma : \overline{K} \to \mathbb{C}$. Given an abelian surface B corresponding to a point $[B]$ on $\mathcal{H}_{\overline{\mathbb{Q}}}$, we use $\sigma([B])$ to denote the corresponding $\mathbb{C}$-point on $\mathcal{H}$ via base change by σ.

[9] We view every Hecke orbit as a horizontal divisor over K, so the arithmetic intersection number is a sum over the finite places of K.

[10] The definition of $T_{\mathfrak{p}}$ depends on the choice of λ if we do not pass to a certain finite quotient of $\mathcal{H}$. However, there are only finitely many choices: let U be the unit group of $\mathcal{O}_F$ and U^+ the subgroup of totally positive units; then the number of choices is $\#U^+/U^2$. Hence we will not specify our choice of λ as it does not affect the arguments in this paper.

[11] The choice of λ determines the polarization.

We set $||\Psi(z)||_{\mathrm{Pet}} = |\Psi(z_1,z_2)(\Im z_1)^{k/2}(\Im z_2)^{k/2}|$, where $z=(z_1,z_2)\in\mathbb{H}^2$. This norm is well-defined outside (the preimage of) $\bigcup_{r\in\mathbb{I}}T(r)$ and invariant under Γ (defined in 2.1.1) and hence we will also view $||\Psi||_{\mathrm{Pet}}$ as a function on $\mathcal{H}_{\mathbb{C}}\backslash\bigcup_{r\in\mathbb{I}}T(r)$. The real analytic function $-\log||\Psi||_{\mathrm{Pet}}$ is a Green function for $\sum_{r\in\mathbb{I}}c_rT(r)$ and endows it with the structure of an arithmetic divisor $\widehat{\sum_{r\in\mathbb{I}}c_rT(r)}$. Note that although the metric $||\cdot||_{\mathrm{Pet}}$ is not defined on the boundary, it only has log-singularity and hence such a Green function is valid in the framework of [BBGK07, §1].

The goal of this section is to show that for most p as in 2.2.1, the archimedean contribution $-\sum_{[B]\in T_{\mathfrak{p}}[A]}\log||\Psi(\sigma([B]))||_{\mathrm{Pet}}$, to the height of $T_{\mathfrak{p}}([\mathcal{A}])$ (with respect to the arithmetic divisor $\widehat{\sum_{r\in\mathbb{I}}c_rT(r)}$), is $o(p\log p)$ as $p\to\infty$. We will discuss this height in detail in §5.1. The equidistribution theorems for Hecke orbits on Shimura varieties reduces this goal to establishing a suitable upper bound (for most p) for $-\log||\Psi(\sigma([B]))||_{\mathrm{Pet}}$ for all $[B]\in T_{\mathfrak{p}}[A]$. We establish this bound by proving that if it is violated for more than one prime p (in a certain range), then $\sigma([\mathcal{A}])$ lies close to a (unique) CM point. The existence of this CM point allows us to show the required bounds for most primes p. The proofs are inspired by Charles' treatment in the case of the modular curve. Recall that all $T(r)$ ($r\in\mathbb{I}$) are compact, so we are able to avoid dealing with estimates around the cusps.

Throughout this section, we view $\mathfrak{a},A,\sigma,\Psi$ as fixed, p is a prime as in 2.2.1 and $N_i(*),C_i(*)$ denote constants only depending on $*$ (here we will not specify the dependence of the constant on $\mathfrak{a},A,\sigma,\Psi$). In particular, if there is no $(*)$, it means an absolute constant (modulo possible dependence on the fixed data). After defining the constants, we may abbreviate $N_i(*),C_i(*)$ as N_i,C_i. Given $\eta\in F$, we use $|\eta|<C$ to mean that for any real embedding $\iota:F\to\mathbb{R}$, the absolute value $|\eta|_\iota<C$. We also use $|\cdot|$ to denote the absolute value on $\mathbb{C}$.

Unless otherwisely specified, we work with the complex analytic topology in this section.

3.1. **An upper bound for values of the Green function on Hecke orbits.**

3.1.1. Let $\mathcal{F}\subset\mathbb{H}^2$ be a fundamental domain for Γ constructed as in [vdG88, I.3] such that the preimage of $\sigma([A])$ in $\mathcal{F}$ lies in the interior[12] of $\mathcal{F}$. Let $\overline{\mathcal{F}}$ denote the closure of $\mathcal{F}$ in $\mathbb{H}^2$ (that is, the cusps of $\Gamma\backslash\mathbb{H}^2$ are not included). For any $\overline{\mathbb{Q}}$-point $[B]$ in $\mathcal{H}$, we use $z(B)=(z_1(B),z_2(B))=(x_1(B)+\sqrt{-1}y_1(B),x_2(B)+\sqrt{-1}y_2(B))$ to denote the preimage of $\sigma([B])$ in $\mathcal{F}$.

Since $T(r)$ is compact for all $r\in\mathbb{I}$, we may choose a compact domain $\Omega\subset\overline{\mathcal{F}}$ such that Ω contains an open neighborhood of the preimage of $\sigma([A])\cup\left(\bigcup_{r\in\mathbb{I}}T(r)\right)$ in $\overline{\mathcal{F}}$. We will fix Ω from now on. Since Ω is compact, there exists $C_0\in\mathbb{R}_{>0}$ such that for any $(z_1,z_2)=(x_1+\sqrt{-1}y_1,x_2+\sqrt{-1}y_2)\in\Omega$, we have $|x_i|<C_0$ and $C_0^{-1}<y_i<C_0$.

Let G be the pull back to $\mathbb{H}^2$ of the Green function $-\log||\Psi||_{\mathrm{Pet}}$ of $\sum c_rT(r)$. There are only finitely many components of the preimage of $T(r)$ in Ω and for each component, we pick (a,b,γ) such that $\begin{pmatrix}a&\gamma\\\gamma'&b\end{pmatrix}\in L$ with $ab-\gamma\gamma'=r\,\mathrm{Nm}\,\mathfrak{a}$ as in 2.1.1 such that this component is defined by $az_1z_2+\gamma z_1+\gamma' z_2+b=0$. We use $\mathcal{M}_{\Omega,r}$ to denote this finite set of (a,b,γ). Then by the definition of the Green function, we have that $G+\sum_{r\in\mathbb{I}}c_r\sum_{(a,b,\gamma)\in\mathcal{M}_{\Omega,r}}\log|az_1z_2+\gamma z_1+\gamma' z_2+b|$ is a real analytic function on $\overline{\mathcal{F}}$.

The goal of this subsection is to show that for most p, one has that

$$-\log|az_1(B)z_2(B)+\gamma z_1(B)+\gamma' z_2(B)+b|\le O(\log p),\forall[B]\in T_{\mathfrak{p}}([A]).$$

Proposition 3.1.2. *Let (a,b,γ) be a fixed triple in $\mathcal{M}_{\Omega,r}$. Given $C_1>20$ and $\epsilon_3>0$, there is an $N_0(\epsilon_3,C_1)>0$ such that for every $N>N_0(\epsilon_3,C_1)$, the number of primes in $[N^{1/2},N]$ for which there exists some $[B]\in T_{\mathfrak{p}}(\sigma([A]))$ such that*

$$|az_1(B)z_2(B)+\gamma z_1(B)+\gamma' z_2(B)+b|<p^{-C_1}$$

is at most $\epsilon_3\#\{\text{primes}\in[N^{1/2},N]\}$.

We extend the idea in [Cha14] of relating bad primes and degrees of homomorphisms between well-chosen CM elliptic curves to the setting of Hilbert modular surfaces by using the theory of special endomorphisms. A point $[B]$ on $\mathcal{H}_{\overline{\mathbb{Q}}}$ is called *special* if there exist $T(n_1)$ and $T(n_2)$, $n_1,n_2\in\mathbb{N}$, $n_1n_2\notin(\mathbb{N})^2$ such that

[12]We may do this because $\mathrm{End}(A)=\mathcal{O}_F$, and thus A doesn't have any automorphisms in addition to $\mathcal{O}_F^\times$.

$[B] \in T(n_1) \cap T(n_2)$. If $[B]$ is special, then B has complex multiplication. We construct a special point $[A_{\mathrm{CM}}]$ on $\mathcal{H}_{\mathbb{C}}$ close to $\sigma([A])$ and show that if some point in $T_{\mathfrak{p}}(\sigma([A]))$ is close to $\mathrm{Div}(\Psi) = \sum_{r\in\mathbb{I}} c_r T(r)$, then A_{CM} has a special endomorphism of certain degree. Proposition 3.1.2 then follows after analysis of the possible degree of special endomorphisms of A_{CM}. In what follows, we will not specify the dependence of the constants C_i and N_i in this subsection on the fixed triple (a, b, γ) and fixed domain Ω.

The following lemma shows that if there exists $[B] \in T_{\mathfrak{p}}(\sigma([A]))$ which is close to $T(r)$, then $\sigma([A])$ is close to $T(pr)$.

Lemma 3.1.3. *If there exists $[B] \in T_{\mathfrak{p}}[A]$ such that $|az_1(B)z_2(B) + \lambda z_1(B) + \lambda' z_2(B) + b| < p^{-C_1}$, then there exist $m \in (D\,\mathrm{Nm}\,\mathfrak{a})\mathbb{Z}, l \in \mathbb{Z}$ and $\eta \in \mathfrak{a}$ such that $ml - \mathrm{Nm}(\eta) = rp\,\mathrm{Nm}\,\mathfrak{a}$ and $|mz_1(A)z_2(A) + \eta z_1(A) + \eta' z_2(A) + l| < p^{-C_1}$. Moreover, we have $|m|, |l|, |\eta| < C_5 p^{1/2}$.*

Proof. As in 2.2.1, we write $\mathfrak{p} = (\lambda)$ and after multiplying λ by an element in $(\mathcal{O}_F^\times)^2$, we may assume that $C_6^{-1}\sqrt{p} < |\lambda| < C_6\sqrt{p}$. We may also assume $z(B) \in \Omega$ (this can be done by letting N be large enough).

Let $U = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \in \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \cdot \Gamma$ be the matrix that maps $z(A)$ to $z(B)$. The set Γ acts on V (in 2.1.1) via $g.M = (g')^t M g$ and this action preserves L. Let $\begin{pmatrix} m & \eta \\ \eta' & l \end{pmatrix} = (U')^t \begin{pmatrix} a & \gamma \\ \gamma' & b \end{pmatrix} U \in L$. Then $ml - \eta\eta' = \det(U')(ab - \gamma\gamma')\det(U) = rp\,\mathrm{Nm}\,\mathfrak{a}$ and

$$mz_1(A)z_2(A) + \eta z_1(A) + \eta' z_2(A) + l = az_1(B)z_2(B) + \lambda z_1(B) + \lambda' z_2(B) + b.$$

This proves the first assertion.

For the second assertion, we first bound $|u_{ij}|$. Consider the real embedding of F corresponding to the first coordinate of $\mathbb{H}^2$ and we will still use u_{ij} to denote its image under this embedding. By definition of U, we have $z_1(B) = \dfrac{u_{11}z_1(A) + u_{12}}{u_{21}z_1(A) + u_{22}}$ and hence $y_1(B) = \dfrac{\lambda y_1(A)}{|u_{21}z_1(A) + u_{22}|^2}$. Since $y_1(B) > C_0^{-1}$ and $y_1(A) < C_0$, we have $|u_{21}z_1(A) + u_{22}|^2 < C_0^2 C_6 \sqrt{p}$. Consider the imaginary part of $u_{21}z_1(A) + u_{22}$ and notice that $y_1(A) > C_0^{-1}$. Thus, we have $|u_{21}| < C_0^2 C_6^{1/2} p^{1/4}$. By considering the real part, we obtain

$$|u_{22}| \le |u_{21}x_1(A) + u_{22}| + |u_{21}x_1(A)| \le |u_{21}z_1(A) + u_{22}| + |u_{21}x_1(A)| \le C_0 C_6^{1/2} p^{1/4} + C_0^3 C_6^{1/2} p^{1/4}.$$

On the other hand, using the bounds of $|u_{21}|, |u_{22}|$, we have

$$|u_{11}| y_1(A) \le |u_{11}z_1(A) + u_{12}| = |(z_1(B)(u_{21}z_1(A) + u_{22}))| \le C_7 p^{1/4}.$$

Hence we obtain

$$|u_{11}| \le C_0 C_7 p^{1/4},\ |u_{12}| \le |u_{11}z_1(A) + u_{12}| + |u_{11}x_1(A)| \le C_7 p^{1/4} + C_0^2 C_7 p^{1/4}.$$

The same argument works for the other embedding of $F \to \mathbb{R}$ by studying $z_2(B), z_2(A)$. The bounds of $|m|, |l|, |\eta|$ follow from the fact that $|u_{ij}|$ is bounded by $O(p^{1/4})$. □

3.1.4. The following lemma shows that if two Hecke orbits of $\sigma([A])$ satisfy the assumption of Lemma 3.1.3, then $\sigma([A])$ is close to a special point on $\mathcal{H}_{\mathbb{C}}$. Recall that for a special point $[B] \in \mathcal{H}(\mathbb{C})$, one defines a quadratic form Q, up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence, as follows (see, for example, [HZ76, 1.1], [vdG88, V.4]). Let $L_{[B]}$ be the sub lattice of L in 2.1.1 such that for any $v \in L_{[B]}$, one has $(z_1(B) \quad 1)v(z_2(B) \quad 1)^t = 0$. The lattice $L_{[B]}$ has rank two and is equipped with a natural orientation. The restriction of the quadratic form on L to the rank two lattice $L_{[B]}$ is positive definite and it coincides with, up to a constant, the quadratic form on the $\mathbb{Z}$-module of special endomorphisms of B. By choosing an oriented basis, one obtains a positive definite integral binary quadratic form Q.

Lemma 3.1.5. *Assume that N is large enough[13] and that for primes $p_1, p_2 \in [N^{1/2}, N], p_1 < p_2$, there exist $[B] \in T_{\mathfrak{p}_1}[A], [B'] \in T_{\mathfrak{p}_2}[A]$ satisfying $|az_1(B)z_2(B) + \lambda z_1(B) + \lambda' z_2(B) + b| < p_1^{-C_1}$ and $|az_1(B')z_2(B') + \lambda z_1(B') + \lambda' z_2(B') + b| < p_2^{-C_1}$. Then there exists a special point $[A_{\mathrm{CM}}]$ on $\mathcal{H}_{\mathbb{C}}$ such that $|z(A) - z(A_{\mathrm{CM}})| < C_8 N^{-\frac{C_1}{2}+2}$ and the integer coefficient binary quadratic form Q_N associated to $[A_{\mathrm{CM}}]$ represents $p_1 r$ and $p_2 r$.*

[13] In the proof, we give a constant $N_3(C_1)$ such that being large enough means $N > N_3$.

Proof. By Lemma 3.1.3, there exist $m_i \in (D\operatorname{Nm}\mathfrak{a})\mathbb{Z}, l_i \in \mathbb{Z}, \eta_i \in \mathfrak{a}$ with $|m_i|, |l_i|, |\eta_i| < C_5 p_i^{1/2}$ such that $m_i l_i - \operatorname{Nm}\eta_i = p_i r \operatorname{Nm}\mathfrak{a}$ and

$$|m_i z_1(A) z_2(A) + \eta_i z_1(A) + \eta_i' z_2(A) + l_i| < p_i^{-C_1}.$$

Since $r = Dq$ for some inert prime q, we have $q||p_i r$. As q is inert, it follows that the norm of any element or ideal in $\mathcal{O}_F$ has to have even q-adic valuation. Therefore, there does not exists $\xi \in F$ such that $\operatorname{Nm}\xi = -p_i r \operatorname{Nm}\mathfrak{a}$. Therefore, $m_i \neq 0$.

We first show that $|\eta_i' + m_i z_1(A)|$ is bounded below by a constant. Indeed, since $m_i \neq 0$, one has $|\eta_i' + m_i z_1(A)| \geq |m_i| y_1(A) \geq C_0^{-1}$.

Therefore,

$$\left| z_2(A) + \frac{\eta_i z_1(A) + l_i}{\eta_i' + m_i z_1(A)} \right| \leq \frac{p_i^{-C_1}}{|\eta_i' + m_i z_1(A)|} \leq C_9 p_i^{-C_1}.$$

Let $f(z)$ be the $\mathcal{O}_F$-coefficient quadratic polynomial in z given by $(\eta_1 z + l_1)(\eta_2' + m_2 z) - (\eta_2 z + l_2)(\eta_1' + m_1 z)$. We first show that the leading coefficient $\eta_1 m_2 - \eta_2 m_1 \neq 0$. If not, then $f(z) = (m_2 l_1 - m_1 l_2) z + l_1 \eta_2' - l_2 \eta_1'$ and $m_2 l_1 - m_1 l_2 \neq 0$ (otherwise, one would have $p_1 = p_2$). In particular, $|m_2 l_1 - m_1 l_2| \geq 1$. Hence $|f(z_1(A))|$ is bounded below by its imaginary part $|m_2 l_1 - m_1 l_2| y_1(A) \geq C_0^{-1}$. On the other hand, since $|\eta_i' + m_i z_1(A)| \leq C_5 p_i^{1/2}(1 + 2C_0)$, we have

$$|f(z_1(A))| \leq \left|(\eta_1' + m_1 z_1(A))(\eta_2' + m_2 z_1(A))\right| \cdot \left| \frac{\eta_1 z_1(A) + l_1}{\eta_1' + m_1 z_1(A)} - \frac{\eta_2 z_1(A) + l_2}{\eta_2' + m_2 z_1(A)} \right| \leq (C_5 + 2C_0 C_5)^2 (p_1 p_2)^{1/2} \cdot (2C_9 p_1^{-C_1}).$$

This leads to a contradiction, when $N^{\frac{C_1}{2} - \frac{3}{4}} > 2C_0(C_5 + 2C_0 C_5)^2 C_9$.

Now we show that f has two complex roots. Let α, β be the two roots of f, then we have $f(z) = (\eta_1 m_2 - \eta_2 m_1)(z - \alpha)(z - \beta)$. Then by assumption, we have

$$2C_9 p_1^{-C_1} > \left| \frac{\eta_1 z_1(A) + l_1}{\eta_1' + m_1 z_1(A)} - \frac{\eta_2 z_1(A) + l_2}{\eta_2' + m_2 z_1(A)} \right| = \left| \frac{(\eta_1 m_2 - \eta_2 m_1)(z_1(A) - \alpha)(z_1(A) - \beta)}{(\eta_1' + m_1 z_1(A))(\eta_2' + m_2 z_1(A))} \right|. \tag{3.1.1}$$

Since $m_i \in \mathbb{Z}$ and $\eta_i \in \mathcal{O}_F$, one has $|\eta_1 m_2 - \eta_2 m_1| \geq |\eta_1' m_2 - \eta_2' m_1|^{-1} \geq (2C_5^2 (p_1 p_2)^{1/2})^{-1}$. Moreover, $|\eta_i' + m_i z_1(A)| \leq C_5 p_i^{1/2}(1 + 2C_0)$. If α, β were real, then $|z_1(A) - \alpha|, |z_1(A) - \beta| \geq C_0^{-1}$ and one gets a contradiction when $N^{\frac{C_1}{2} - \frac{3}{2}} > 4C_0^2 C_5^4 C_9 (1 + 2C_0)^2$. Then we may assume that α has positive imaginary part.

Consider the point $\left(\alpha, -\frac{\eta_1 \alpha + l_1}{\eta_1' + m_1 \alpha}\right) \in \mathbb{H}^2$. We now show that this point is close to $z(A)$. We use eqn. (3.1.1) to study the first coordinate. Since β has negative imaginary part, $|z_1(A) - \beta| \geq C_0^{-1}$. Since $p_2 \leq p_1^2$, the above discussion shows that

$$|z_1(A) - \alpha| \leq 4C_0 C_5^4 C_9 (1 + 2C_0)^2 p_1^{-C_1 + 3}.$$

Moreover, for the second coordinate,

$$\begin{aligned} \left| z_2(A) - \left(-\frac{\eta_1 \alpha + l_1}{\eta_1' + m_1 \alpha} \right) \right| &\leq \left| \frac{\eta_1 z_1(A) + l_1}{\eta_1' + m_1 z_1(A)} - \frac{\eta_1 \alpha + l_1}{\eta_1' + m_1 \alpha} \right| + \left| z_2(A) + \frac{\eta_1 z_1(A) + l_1}{\eta_1' + m_1 z_1(A)} \right| \\ &\leq \frac{(l_1 m_1 - \eta_1 \eta_1')|z_1(A) - \alpha|}{|(\eta_1' + m_1 z_1(A))(\eta_1' + m_1 \alpha)|} + C_9 p_1^{-C_1} \\ &\leq 4C_0^2 C_5^4 C_9 (1 + 2C_0)^2 C_{10} r (\operatorname{Nm}\mathfrak{a}) p_1^{-C_1 + 4}, \end{aligned}$$

where C_{10}^{-1} is the lower bound of $|\eta_1' + m_1 \alpha|$ given by $C_0^{-1} - 4C_0 C_5^4 C_9 (1 + 2C_0)^2 p_1^{-C_1 + 3}$.

Since $\Omega \cap \mathcal{F}$ contains a (fixed) open neighborhood of $z(A)$, then the point $\left(\alpha, -\frac{\eta_1 \alpha + l_1}{\eta_1' + m_1 \alpha}\right)$ lies in $\Omega \cap \mathcal{F}$ when N is large enough.

We define A_{CM} to be the abelian surface such that $z(A_{\mathrm{CM}}) = \left(\alpha, -\frac{\eta_1 \alpha + l_1}{\eta_1' + m_1 \alpha}\right)$. Then A_{CM} lies on both $T(p_1 r)$ and $T(p_2 r)$ (so $[A_{\mathrm{CM}}]$ is a special point). In other words, the integer coefficient binary quadratic form associated to $[A_{\mathrm{CM}}]$ represents $p_1 r$ and $p_2 r$.

Moreover, we may take $C_8 = 8C_0 C_5^4 C_9 (1 + 2C_0)^2 C_0^2 r (\operatorname{Nm}\mathfrak{a})$ to get the desired inequality when N is large enough. (Note that we take C_{10} to be C_0 here since we have proved that $z(A_{\mathrm{CM}}) \in \Omega$.) $\square$

The following lemma shows that the special point $[A_{\mathrm{CM}}]$ constructed above is unique when N is large enough.

Lemma 3.1.6. *Assume that N is large enough*[14] *and that for $p_i \in [N^{1/2}, N]$ with $i = 1, 2, 3, 4$, $p_1 \neq p_2$, $p_3 \neq p_4$, there exists $[B_i] \in T_{\mathfrak{p}_i}([A])$ such that $|az_1(B_i)z_2(B_i) + \lambda z_1(B_i) + \lambda' z_2(B_i) + b| < p_i^{-C_1}$. Let $[A_1], [A_2]$ be two special points constructed as in Lemma 3.1.5 by using the assumption on p_1, p_2 and p_3, p_4. Then $[A_1] = [A_2]$. More precisely, if $|z(A_1) - z(A_2)| \leq 2C_8 N^{-\frac{C_1}{2}+2}$, then $z(A_1) = z(A_2)$.*

Proof. Assume for contradiction that $z(A_1) \neq z(A_2)$. Without loss of generality, we may assume that $z_1(A_1) \neq z_1(A_2)$; otherwise, we may redo the proof assuming $z_2(A_1) \neq z_2(A_2)$ and the same estimate works.

Let $f_1(z), f_2(z) \in \mathcal{O}_F[z]$ be the quadratic equations defining $z_1(A_1), z_1(A_2)$ in the proof of Lemma 3.1.5. Here F is naturally a subfield of $\mathbb{R}$ with respect to the embedding of first copy of $\mathbb{H}$ in $\mathbb{H}^2$. Let $\alpha_1, \alpha_2 \in F$ (resp. $\beta_1, \beta_2 \in F$) be the leading coefficients (resp. the coefficients of z) of f_1, f_2 and $-\Delta_1, -\Delta_2$ the discriminants. Since f_i has two complex roots for both embeddings of F, one has that Δ_i is totally real. We may assume both α_1, α_2 are positive with respect to this embedding $F \subset \mathbb{R}$.

We discuss two cases: (1) $y_1(A_1) \neq y_1(A_2)$ and (2) $x_1(A_1) \neq x_1(A_2)$. Since $z_1(A_1) \neq z_1(A_2)$, at least one of the above situations happen.

In case (1), we have $\frac{\sqrt{\Delta_1}}{2\alpha_1} = y_1(A_1) \neq y_1(A_2) = \frac{\sqrt{\Delta_2}}{2\alpha_2}$. By the definition of f_i and Lemma 3.1.3, we have $|\alpha_i|, |\alpha_i'| \leq 2C_5^2 N$ and $|\Delta_i|, |\Delta_i'| \leq C_{11} N^2$. Moreover, since $\alpha_i, \Delta_i \in \mathcal{O}_F$, we have the nonzero $|\operatorname{Nm}(\alpha_2^2\Delta_1 - \alpha_1^2\Delta_2)| \geq 1$ and hence

$$|\alpha_2^2\Delta_1 - \alpha_1^2\Delta_2| \geq |(\alpha_2')^2\Delta_1' - (\alpha_1')^2\Delta_2'|^{-1} \geq (8C_5^4 C_{11})^{-1} N^{-4}.$$

Putting these inequalities together, we obtain

$$\begin{aligned} |z(A_1) - z(A_2)| &\geq |z_1(A_1) - z_1(A_2)| \geq |y_1(A_1) - y_1(A_2)| = \Big|\frac{\sqrt{\Delta_1}}{2\alpha_1} - \frac{\sqrt{\Delta_2}}{2\alpha_2}\Big| \\ &= \frac{|\alpha_2^2\Delta_1 - \alpha_1^2\Delta_2|}{2\alpha_1\alpha_2(\alpha_1\sqrt{\Delta_2} + \alpha_2\sqrt{\Delta_1})} \geq C_{12} N^{-8}. \end{aligned}$$

This contradicts our assumption when $N^{(C_1/2)-10} > 2C_8 C_{12}^{-1}$.

In case (2), we have $\frac{-\beta_1}{2\alpha_1} = x_1(A_1) \neq x_1(A_2) = \frac{-\beta_2}{2\alpha_2}$. The argument is similar to case (1). By the definition of f_i and Lemma 3.1.3, we have $|\beta_i|, |\beta_i'| \leq 4C_5^2 N$ and then $|\alpha_2\beta_1 - \alpha_1\beta_2| \geq |\alpha_2'\beta_1' - \alpha_1'\beta_2'|^{-1} \geq (16C_5^4)^{-1} N^{-2}$. We conclude that

$$|z(A_1) - z(A_2)| \geq |x_1(A_1) - x_2(A_2)| = \left|\frac{\beta_1}{2\alpha_1} - \frac{\beta_2}{2\alpha_2}\right| = \left|\frac{\beta_1\alpha_2 - \beta_2\alpha_1}{2\alpha_1\alpha_2}\right| \geq C_{13} N^{-4}.$$

This contradicts our assumption when $N^{(C_1/2)-6} > 2C_8 C_{13}^{-1}$. □

Corollary 3.1.7. *Assume that N is large enough as above and that A satisfies the assumption in Lemma 3.1.5. For any $p_3 \in [\sqrt{N}, N]$ such that there exists $[B''] \in T_{\mathfrak{p}_3}[A]$ satisfying $|az_1(B'')z_2(B'') + \lambda z_1(B'') + \lambda' z_2(B'') + b| < p_3^{-C_1}$, the quadratic form Q_N in Lemma 3.1.5 represents $p_3 r$.*

Proof. We may assume that $p_3 \neq p_1$. By Lemma 3.1.5, we construct special points $[A_1]$ by using p_1, p_2 and $[A_2]$ by using p_1, p_3. By Lemma 3.1.6, we have $[A_1] = [A_2]$ and hence they have the same quadratic form Q_N. Since $[A_2]$ lies on $T(p_3 r)$, then Q_N represents $p_3 r$. □

Lemma 3.1.8. *Fix $C_1 > 20$ and let Δ_N denote the discriminant of Q_N in Lemma 3.1.5. As $N \to \infty$, we have $|\Delta_N| \to \infty$.*

Proof. Fix a bound X of $|\Delta_N|$, then there are only finitely many equivalent classes of integral binary quadratic forms of discriminant $\leq X$. For each class, [HZ76, Thm. 1][15] shows that there are only finitely many special points corresponding to the given class of quadratic forms. As $N \to \infty$, the CM approximation $[A_{\mathrm{CM},N}]$ is closer to $\sigma([A])$ and hence $|\Delta_N|$ cannot be bounded. □

[14] In the proof, we give a constant $N_4(C_1)$ such that being large enough means $N > N_4$.

[15] Although in [HZ76], they assume that D is a prime, their method still works in general. See for example [vdG88, V.6].

Proof of Proposition 3.1.2. Let p_1, p_2 be the smallest primes in $[\sqrt{N}, N]$ such that there exists $B \in T_{\mathfrak{p}_i} A$ such that $|az_1(B)z_2(B)+\lambda z_1(B)+\lambda' z_2(B)+b| < p_i^{-C_1}$. Then by Lemma 3.1.5 and Corollary 3.1.7, we obtain a quadratic form Q_N associated to a special point $[A_{\mathrm{CM},N}]$ which represents $p_3 r$ for any prime p_3 in $[\sqrt{N}, N]$ which satisfies the condition that there exists $[B''] \in T_{\mathfrak{p}_3}[A]$ satisfying $|az_1(B'')z_2(B'')+\lambda z_1(B'')+\lambda' z_2(B'')+b| < p_3^{-C_1}$. There exist finitely many (and this number is bounded by $2^{d(r)}$, where $d(r)$ is the number of factors of r) positive definite integral binary quadratic forms Q'_N such that a prime pr is represented by Q_N only if p is represented by some Q'_N. One also has that the absolute value of the discriminant Δ'_N of Q'_N is at least $|\Delta_N|/r^2$.

The result now follows from the following claim.

Claim 3.1.9. We have $\dfrac{\{\text{prime } p \in [N^{1/2}, N], p \text{ is represented by } Q'_N\}}{\{p \in [N^{1/2}, N]\}} \to 0$ as $N \to \infty$.

Proof. When $|\Delta'_N| \leq (\log N)^4$, [TZ16, Corollary 1.3] shows that

$$\{\text{prime } p \in [N^{1/2}, N], p \text{ is represented by } Q'_N\} \ll \frac{\mathrm{Li}(N)}{h_N},$$

where h_N is the number of $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of primitive positive definite integral binary quadratic forms of discriminant Δ'_N. Since $\Delta'_N \to -\infty$ by Lemma 3.1.8, one has $h_N \to \infty$. When $|\Delta'_N| > (\log N)^4$, [Cha14, Lemma 5.2] shows that

$$\{\text{integer } n \in [N^{1/2}, N], n \text{ is represented by } Q'_N\} \leq 1 + 4\sqrt{2N} + 8N/|\Delta'_N|^{1/2} = O\left(\frac{N}{(\log N)^2}\right).$$

The claim follows by putting these two cases together. □

□

3.2. **From equidistribution to an upper bound of archimedean contribution.** This is the main theorem of this section. We use Proposition 3.1.2 and equidistribution theorem for Hecke orbits to show that for most p, the archimedean contribution in the height of $T_{\mathfrak{p}}([\mathcal{A}])$ is $o(p \log p)$.

Theorem 3.2.1. *For any $\epsilon_1, \epsilon_2 > 0$, there is an $N(\epsilon_1, \epsilon_2) > 0$ such that for every $N > N(\epsilon_1, \epsilon_2)$, the number of primes in the interval $[N^{1/2}, N]$ for which*

$$-\sum_{[B] \in T_{\mathfrak{p}}[A]} \log ||\Psi(\sigma([B]))||_{\mathrm{Pet}} \geq \epsilon_1 p \log p$$

is at most $\epsilon_2 \#\{\ell \in [N^{1/2}, N] \text{ prime}\}$.

Proof. Notation as in 3.1.1. We first show that a fixed triple (a, b, γ),

$$-\sum_{B \in T_{\mathfrak{p}} A} \log |az_1(B)z_2(B) + \gamma z_1(B) + \gamma' z_2(B) + b| = o(p \log p).$$

For any $\epsilon > 0$, by the equidistribution theorem of Hecke orbits (see for example [COU01]), there exist constants $N_1(\epsilon, C_1) > 0$ and $C_3(\epsilon, C_1) < 0$ such that for any $p > N_1$,

$$\#\{[B] \in T_{\mathfrak{p}}[A] : \log |az_1(B)z_2(B) + \gamma z_1(B) + \gamma' z_2(B) + b| < C_3\} < \epsilon p / C_1.$$

Let $\mathbb{I}' = \{r \in \mathbb{I} \mid c(r) > 0\}$ and $M = \sum_{r \in \mathbb{I}'} \#\mathcal{M}_{\Omega,r}$. Taking $\epsilon_3 = \epsilon_2/M$ and applying Proposition 3.1.2 for all triples $(a, b, \gamma) \in \bigcup_{r:c(r)>0} \mathcal{M}_{\Omega,r}$, we have that, for every $N > N_2(\epsilon_2, C_1) := \max_{(a,b,\gamma)}\{N_0(a, b, \gamma, \epsilon_3, C_1), N_1^2\}$,

$$-\sum_{[B] \in T_{\mathfrak{p}}[A]} \log |az_1(B)z_2(B) + \gamma z_1(B) + \gamma' z_2(B) + b| < -(p+1)C_3 + (\epsilon p / C_1) \cdot C_1 \log p$$

holds for primes $p \in [N^{1/2}, N]$ outside a set $\mathbb{B}_N$ of density ϵ_2 (this set is the union of the exceptional sets for all $(a, b, \gamma) \in \cup_{r:c(r)>0} \mathcal{M}_{\Omega,r}$).

Let ϕ be a smooth function which is 1 in Ω with compact support in $\overline{\mathcal{F}}$. Then by 3.1.1, the function $f = G + \sum_{r \in \mathbb{I}} c(r) \sum_{(a,b,\gamma) \in \mathcal{M}_{\Omega,r}} \phi(z) \log |az_1 z_2 + \gamma z_1 + \gamma' z_2 + b|$ is smooth on $\overline{\mathcal{F}}$. Since Ψ does not have poles or zeros along the boundary of $\overline{\mathcal{H}}_{\mathbb{C}}^{\mathrm{tor}}$ and the Petersson metric admits logarithm singularity along the

boundary, G and hence f go to $-\infty$ as (z_1, z_2) approaches the cusps. Therefore, f is bounded above on $\overline{\mathcal{F}}$. On the other hand, since $\phi(z)$ has compact support, $\phi(z)\log|az_1z_2+\gamma z_1+\gamma' z_2+b|$ is also bounded above. Therefore, since $\#T_{\mathfrak{p}}[A]=p+1$, we have

$$\sum_{[B]\in T_{\mathfrak{p}}[A]}\big(G(z(B))+\sum_{r\in\mathbb{I}'}c(r)\sum_{(a,b,\gamma)\in\mathcal{M}_{\Omega,r}}\phi(z(B))\log|az_1(B)z_2(B)+\gamma z_1(B)+\gamma' z_2(B)+b|\big)<C_2\cdot(p+1).$$

Take $\epsilon=\epsilon_1/(2M')$, where $M'=\sum_{r\in\mathbb{I}'}c(r)\#\mathcal{M}_{\Omega,r}$, then we have, for $N>N_2$, for $p\in[N^{1/2},N]\backslash\mathbb{B}_N$,

$$\sum_{[B]\in T_{\mathfrak{p}}[A]}G(z(B))<C_4(p+1)+(\epsilon_1 p\log p)/2.$$

Then by taking $N(\epsilon_1,\epsilon_2)>N_2$ large enough so that $C_4<(\epsilon_1\log N(\epsilon_1,\epsilon_2))/4$, the theorem follows. □

4. Special endomorphisms and contributions at finite places

We first establish notation for this section. Let ℓ denote a prime, and $B, B', \mathcal{B}, \mathcal{B}'$ denote abelian surfaces with $\mathcal{O}_F$-multiplication and $\mathfrak{a}$-polarization. We fix a finite place v of a number field K over ℓ. Recall that e_v is the ramification degree of K at v. These abelian surfaces may be defined over $K, \mathcal{O}_K$, $\mathcal{O}_{K_v}$, or $\mathcal{O}_{K_v}/v^n$. Recall that we use $\mathcal{B}, \mathcal{B}'$ to denote abelian surfaces defined over $\mathcal{O}_{K_v}$ and we use $\mathcal{B}_{v,n}, \mathcal{B}'_{v,n}$ to denote their reduction modulo v^n. We will use $\mathcal{B}_{\ell,n}$ and $\mathcal{B}'_{\ell,n}$ to denote surfaces over $\mathcal{O}_{K_v}/\ell^n$ (note that the abelian varieties are defined modulo ℓ^n, not v^n) which do not *a priori* come with lifts to $\mathcal{O}_{K_v}$. Let $M_{v,n}$ and $M_{\ell,n}$ denote the module of special endomorphisms of $\mathcal{B}_{v,n}$ and $\mathcal{B}_{\ell,n}$ respectively. Finally, we let $\Lambda_v=\mathrm{End}(\mathcal{B}[\ell^\infty])\cap(M_{v,1}\otimes\mathbb{Z}_\ell)$, where the intersection takes place in $\mathrm{End}(\mathcal{B}_{v,1})\otimes\mathbb{Z}_\ell=\mathrm{End}(\mathcal{B}_{v,1}[\ell^\infty])$. We call the $\mathbb{Z}_\ell$-module Λ_v the set of special endomorphisms of $\mathcal{B}[\ell^\infty]$

In this section, we will bound the local intersection multiplicities of $(T_{\mathfrak{p}}([\mathcal{A}]),\mathcal{T}(r))$ at non-archimedean places for $r\in\mathbb{I}$, where $\mathbb{I}$ is a fixed finite set such that $D|r$ and $\mathcal{T}(r)$ is compact in $\mathcal{H}$ for all $r\in\mathbb{I}$. The choice of $\mathbb{I}$ will be dictated by Lemma 5.1.1. We will relate the local intersection multiplicity to special endomorphisms of $\mathcal{A}_{v,n}$. The main step involved in deducing the required bounds is to bound the number of special endomorphisms of $\mathcal{A}_{v,n}$. We prove that the nested sequence of modules $M_{v,n}$ shrink sufficiently rapidly. The precise statement is in Theorem 4.1.1. Together with geometry-of-numbers arguments, this result allows us to bound the number of special endomorphisms of $\mathcal{A}_{v,n}$.

4.1. Deformation theory.

The following result is crucial to bounding the local intersections:

Theorem 4.1.1. *There exists a positive integer n_0 such that $M_{v,n_0'+ke_v}=(\Lambda_v+\ell^k M_{v,n_0'}\otimes\mathbb{Z}_\ell)\cap M_{v,n_0}$ where $n_0'\geq n_0$ and k is allowed to be any positive integer.*[16]

We will need the following result on homomorphisms between abelian surfaces:

Lemma 4.1.2. *Let $\alpha\in\mathrm{Hom}(\mathcal{B}_{\ell,n-1},\mathcal{B}'_{\ell,n-1})$ for any $n\geq 3$.*

1. *The homomorphism $\ell\alpha$ lifts uniquely to $\mathrm{Hom}(\mathcal{B}_{\ell,n},\mathcal{B}'_{\ell,n})$, where $\mathcal{B}_{\ell,n}$ and $\mathcal{B}'_{\ell,n}$ are lifts of $\mathcal{B}_{\ell,n-1}$ and $\mathcal{B}'_{\ell,n}$.*
2. *If $\ell\alpha$ lifts to $\mathrm{Hom}(\mathcal{B}_{\ell,n+1},\mathcal{B}'_{\ell,n+1})$, then α lifts to $\mathrm{Hom}(\mathcal{B}_{\ell,n},\mathcal{B}'_{\ell,n})$. Here, $\mathcal{B}_{\ell,n+1}$ and $\mathcal{B}'_{\ell,n+1}$ are lifts of $\mathcal{B}_{\ell,n}$ and $\mathcal{B}'_{\ell,n}$.*

Let $\mathscr{G}_i=\mathcal{B}_{\ell,i}[\ell^\infty]$, and $\mathscr{G}'_i=\mathcal{B}'_{\ell,i}[\ell^\infty]$. By the Serre–Tate lifting theorem, it suffices to prove the analogous result for ℓ-divisible groups. This is a straightforward application of Grothendieck–Messing theory. Before proceeding to the proof, we recall some facts from Grothendieck–Messing theory ([Mes72] contains every result that we need). All the reduction maps between the $\mathcal{O}_{K_v}/\ell^i$ for $i=n,n\pm1$ are canonically equipped with nilpotent divided powers (in fact, as $n>2$, all the ideals in play are square-zero). Let $\mathbb{D}$ and $\mathbb{D}'$ denote the Dieudonné-crystals associated to $\mathscr{G}_{n-1}$ and $\mathscr{G}'_{n-1}$ (see [Mes72, §2.5 of Chapter IV]). Any homomorphism between $\mathscr{G}_{n-1}$ and $\mathscr{G}'_{n-1}$ canonically induces a map of crystals $\mathbb{D}\to\mathbb{D}'$.

Let D_i and D'_i denote $\mathbb{D}$ and $\mathbb{D}'$ evaluated at $\mathcal{O}_{K_v}/\ell^i$ for $i=n,n\pm1$ (these are free $\mathcal{O}_{K_v}/\ell^i$-modules whose ranks equal the heights of $\mathscr{G}$ and $\mathscr{G}'$). Grothendieck–Messing theory associates canonical filtrations

[16] In an earlier version of the paper, we implicitly assumed that $\Lambda_v=0$ (this assumption simplified the geometry-of-numbers arguments), and we are very grateful to Davesh Maulik for pointing this out to us.

$F_i \subset D_i$ and $F'_i \subset D'_i$ to the groups $\mathscr{G}_i$ and $\mathscr{G}'_i$ for $i = n, n \pm 1$. Note that the F_{n+1} reduces to F_n and F_{n-1} under the canonical quotient maps (the analogous statement holds for F'_{n+1}). The filtrations are direct summands of the crystals evaluated at the $\mathcal{O}_{K_v}/\ell^i$. Suppose that $W'_{n+1} \subset D'_{n+1}$ is some submodule such that $D'_{n+1} = F'_{n+1} \oplus W'_{n+1}$. Let $W'_n \subset D'_n$ denote the mod-ℓ^n reduction of W'_{n+1}. Clearly, $W'_n \oplus F'_n = D'_n$. By [Mes72, Theorem 1.6 of Chapter V], a homomorphism between $\mathscr{G}_{n-1}$ and $\mathscr{G}'_{n-1}$ lifts to $\mathrm{Hom}(\mathscr{G}_i, \mathscr{G}'_i)$ (for $i = n, n+1$) if and only if the associated map of crystals evaluated at $\mathcal{O}_{K_v}/\ell^i$ maps the filtration F_i to F'_i. Let $\alpha_i : D_i \to D'_i$ ($i = n, n \pm 1$) denote the maps induced by α. We now proceed to the proofs of the two statements.

Proof of Lemma 4.1.2.

(1) Let $v_n \in F_n \subset D_n$, whose image in F_{n-1} is denoted by v_{n-1}. It suffices to prove that $\ell\alpha_n(v_n) \in F'_n$. Let $\alpha_n(v_n) = v'_n + w'_n$, where $v'_n \in F'_n$ and $w'_n \in W'_n$. As $\alpha_{n-1}(F_{n-1}) \subset F'_{n-1}$, we have w'_n modulo ℓ^{n-1} is zero. It follows that $\ell w'_n = 0$. Therefore, $\ell\alpha_n(v_n) = \ell v'_n \in F'_n$. Thus $\ell\alpha_n$ preserves filtrations, as required.

(2) As above, let $v_n \in F_n$. Let $v_{n+1} \in F_{n+1}$, whose mod-ℓ^n reduction is v_n. Suppose that $\alpha_{n+1}(v_{n+1}) = v'_{n+1} + w'_{n+1}$, where $v'_{n+1} \in F'_{n+1}$ and $w'_{n+1} \in W'_{n+1}$. As $\ell\alpha$ lifts to $\mathrm{Hom}(\mathscr{G}_{n+1}, \mathscr{G}'_{n+1})$, it follows that $\ell w'_{n+1} = 0$. Therefore, $w'_{n+1} = 0$ modulo ℓ^n. It follows that $\alpha_{n+1}(v_{n+1})$ modulo ℓ^n - which equals $\alpha_n(v_n)$ - is an element of F'_n.

□

Lemma 4.1.3. *We have* $\cap_{n=1}^{\infty}(M_{v,n} \otimes \mathbb{Z}_\ell) = \Lambda_v$.

Proof. Since $\mathrm{End}(\mathcal{B}_{v,1}[\ell^\infty]) = \mathrm{End}(\mathcal{B}_{v,1}) \otimes \mathbb{Z}_\ell$, then by Serre–Tate theory, $M_{v,n} \otimes \mathbb{Z}_\ell = \mathrm{End}(B_{v,n}[\ell^\infty]) \cap (M_{v,1} \otimes \mathbb{Z}_\ell)$. Therefore, $\cap_{n=1}^{\infty}(M_{v,n} \otimes \mathbb{Z}_\ell)$ is the lattice of special endomorphism of the formal ℓ-divisible group $\mathcal{B}[\ell^\infty]$ over $\mathrm{Spf}\mathcal{O}_{K_v}$. Note that the category of ℓ-divisible groups over $\mathrm{Spf}\mathcal{O}_{K_v}$ is equivalent to the category of ℓ-divisible groups over $\mathcal{O}_{K_v}$. Hence we obtain the desired assertion. □

We now prove Theorem 4.1.1

Proof of Theorem 4.1.1. For ease of notation, denote by $\Lambda_{v,i}$ the $\mathbb{Z}_\ell$-module $M_{v,i} \otimes \mathbb{Z}_\ell$. By the Serre–Tate theorem, it suffices to prove the existence of n_0 such that $\Lambda_{v,n'_0+ke_v} = \Lambda_v + \ell^k\Lambda_{v,n'_0}$ for $n'_0 \geq n_0$. First, note that Lemma 4.1.2 implies that $\Lambda_v \subset \Lambda_{v,2e_v}$ is co-torsion free. Let $\Lambda' \subset \Lambda_{v,2e_v}$ denote a direct summand of Λ_v.

Consider the following statement:

Claim 4.1.4. We have that $\Lambda' \cap \Lambda_{v,n+e_v} \subset \ell(\Lambda' \cap \Lambda_{v,n})$ for large enough n.

Proof of claim. Fix any $n' > 2e_v$. As the $\mathbb{Z}_\ell$-module of special endomorphisms of $\mathcal{B}[\ell^\infty]$ equals Λ_v, it follows that $\bigcap_n(\Lambda' \cap \Lambda_{v,n}) = 0$. Therefore $\Lambda' \cap \Lambda_{v,n'+ke_v} \subset \ell(\Lambda' \cap \Lambda_{v,n'})$ for large enough k. We now prove by contradiction that $\Lambda' \cap \Lambda_{v,n'+(k+1)e_v} \subset \ell(\Lambda' \cap \Lambda_{v,n'+ke_v})$ for such k.

Assume that there exists a special endomorphism $\alpha \in (\Lambda' \cap \Lambda_{v,n'+(k+1)e_v}) \backslash \ell(\Lambda' \cap \Lambda_{v,n'+ke_v})$. If $\alpha \in \ell(\Lambda' \cap \Lambda_{v,n'+(k-1)e_v})$, we write $\alpha = \ell\beta$, where $\beta \in \Lambda' \cap \Lambda_{v,n'+(k-1)e_v}$. By assumption, $\ell\beta \in \Lambda' \cap \Lambda_{v,n'+(k+1)e_v}$ and then by Lemma 4.1.2, $\beta \in \Lambda' \cap \Lambda_{v,n'+ke_v}$. This contradicts that $\alpha \notin \ell(\Lambda' \cap \Lambda_{v,n'+ke_v})$ and hence we have shown that $\alpha \notin \ell(\Lambda' \cap \Lambda_{v,n'+(k-1)e_v})$. By iterating this argument, it follows that $\alpha \notin \ell(\Lambda' \cap \Lambda_{v,n'})$, which is a contradiction. □

We now prove the theorem. Choose n_0 such that the above claim holds for every $n \geq n_0$. For any $n'_0 \geq n_0$, the inclusion $\Lambda_{v,n'_0+ke_v} \supset \Lambda_v + \ell^k\Lambda_{v,n'_0}$ follows from Lemma 4.1.2. As for the other inclusion, let $\Lambda'_k = \Lambda' \cap \Lambda_{v,n'_0+ke_v}$. It is clear that $\Lambda_{v,n'_0+ke_v} = \Lambda_v \oplus \Lambda'_k$, and it suffices to prove that $\Lambda'_k \subset \ell^k\Lambda'_0$. We induct on k by assuming this holds, and then proving that $\Lambda'_{k+1} \subset \ell\Lambda'_k$. However, this follows from the claim and the fact that Λ'_k is co-torsion free in Λ_{v,n'_0+ke_v}. □

4.2. **Geometry of numbers and applications to counting special endomorphisms.** For an m-dimensional lattice M with a positive definite quadratic form Q, let $\mu_1(M) \leq \mu_2(M) \ldots \leq \mu_m(M)$ denote the successive minima of M (see [EK95, Definition 2.2] for the definition of the term successive minima). We will need the following lemma due to Schmidt:

Lemma 4.2.1. *The following estimate holds:* $\#\{s \in M \mid Q(s) \leq N\} = O\left(\sum_{j=0}^{m} \frac{N^{j/2}}{\mu_1(M)\dots\mu_j(M)}\right)$, *where the implied constant depends only on* m.

Proof. Equations (5) and (6) on page 518 of [EK95] together with Lemma 2.4 of *loc. cit.* implies the stated result (the authors refer to [Sch68] for a proof of Lemma 2.4). □

We now prove an elementary lemma that will allow us to use Lemma 4.2.1 to bound special endomorphisms of $\mathcal{B}_{v,n}$. We refer to the beginning of this section for notation.

Lemma 4.2.2. *Let* m *be the* $\mathbb{Z}$*-rank of* $M_{v,1}$ *and* m' *be the* $\mathbb{Z}_\ell$*-rank of* Λ_v. *Then* $\prod_{i=1}^{j} \mu_i(M_{v,n}) \gg \ell^{n(j-m')/e_v}$.

Proof. We may assume that $m' < j$. It suffices to prove that $\prod_{i=1}^{j} \mu_i(M_{v,n_0+ke_v}) \gg \ell^{k(j-m')}$.

For a lattice M, let $d(M)$ denote the square root of its discriminant. Theorem 4.1.1 implies that $d(M_{v,n_0+ke_v}) \geq \ell^{k(m-m')}$. Thus,

$$\prod_{i=1}^{m} \mu_i(M_{v,n_0+ke_v}) \gg \ell^{k(m-m')} \tag{4.2.1}$$

by [EK95, Eqn. (5),(6)]. This is the desired result for $j = m$.

Moreover, if $M \subset M'$ are lattices, then $\mu_i(M) \geq \mu_i(M')$. Therefore, Theorem 4.1.1 implies that $\mu_i(M_{v,n_0+ke_v}) \leq \mu_i(\ell^k M_{v,n_0}) = \ell^k \mu_i(M_{v,n_0}) \ll \ell^k$. The lemma follows from multiplying (4.2.1) with the inequality $\prod_{i=j+1}^{m} \mu_i(M_{v,n_0+ke_v})^{-1} \gg \ell^{k(j-m)}$. □

Lemmas 4.2.1 and 4.2.2 immediately yield the following corollary:

Corollary 4.2.3. *Suppose that the* $\mathbb{Z}_\ell$*-rank of* Λ_v *is* ≤ 1 *and the rank of* $M_{v,1}$ *is* m. *Then*

$$\#\{s \in M_{v,n} \mid Q(s) \leq N\} = O\Big(N^{1/2} + \sum_{j=2}^{m} \frac{N^{j/2}}{\ell^{(j-1)n/e_v}}\Big).$$

4.3. **Proof of the non-archimedean local results.** In what follows, we consider $T_{\mathfrak{p}}$ as in 2.2.1 with $p \neq \ell$ and recall that we write $(p) = \mathfrak{p}\mathfrak{p}' \subset \mathcal{O}_F$. We always use N to denote a large enough integer.

Lemma 4.3.1. *Over* $\mathbb{Z}[1/pr]$, *we have* $T_{\mathfrak{p}}\mathcal{T}(r) = \mathcal{T}(pr) = T_{\mathfrak{p}'}\mathcal{T}(r)$. *Moreover, for any* n, *if there exists* $[\mathcal{B}] \in T_{\mathfrak{p}}[\mathcal{A}]$ *such that* $[\mathcal{B}_{v,n}] \in \mathcal{T}(r)$, *then* $[\mathcal{A}_{v,n}] \in \mathcal{T}(pr)$.

Proof. Checking at the level of complex points, we have $T_{\mathfrak{p}}T(r) = T(pr) = T_{\mathfrak{p}'}\mathcal{T}(r)$. By definition, $\mathcal{T}(m)$ is the Zariski closure of $T(m)$ and hence $T_{\mathfrak{p}}\mathcal{T}(r) = \mathcal{T}(pr)$. The second assertion then follows from the étaleness of Hecke orbits. □

For the rest of this section, we assume $\mathrm{End}(A_{\overline{\mathbb{Q}}}) = \mathcal{O}_F$ and $\mathcal{A}$ has good reduction at v. Further, the *norm* of a special endomorphism s denotes the integer $Q(s)$.

The following lemma is well-known and follows directly from the crystalline realization of the module of special endomorphisms. We record a proof here for completeness.

Lemma 4.3.2. *Let* m *be the* $\mathbb{Z}$*-rank of* $M_{v,1}$ *and* m' *be the* $\mathbb{Z}_\ell$*-rank of* Λ_v. *Then* $m \leq 4$ *and* $m' \leq 2$.

Proof. Since $s^* = s$ for any $s \in M_{v,1}$ (resp. Λ_v), we have that $\Lambda_v \subset M_{v,1} \otimes \mathbb{Z}_\ell \subset H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))$. On the other hand, as $\mathcal{O}_F \subset \mathrm{End}(\mathcal{B}_{v,1})$ is stable under the Rosati involution, we have a natural embedding $\mathcal{O}_F \subset H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))$. Since $H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))[1/\ell]$ is a $W(\bar{\mathbb{F}}_\ell))[1/\ell]$-vector space of dimension 6, the Frobenius invariant part $H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))[1/\ell]^{\varphi=1}$ is a $\mathbb{Q}_\ell$-vector space of dimension at most 6.

As $s \circ f = f' \circ s$ for any $s \in M_{v,1} \otimes \mathbb{Q}_\ell$ and $f \in F$, it follows that $\mathcal{O}_F \otimes \mathbb{Q}_\ell \cap M_{v,1} \otimes \mathbb{Q}_\ell = 0$ in $H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))[1/\ell]^{\varphi=1}$. Since $\mathcal{O}_F \otimes \mathbb{Q}_\ell$ has dimension 2, we have that $M_{v,1} \otimes \mathbb{Q}_\ell$ is at most of dimension 4 and hence $m \leq 4$.

On the other hand, the de Rham cohomology of $\mathcal{B}$ induces a (decreasing) Hodge filtration $\mathrm{Fil}^\bullet$ on $H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell)) \otimes \bar{\mathbb{Q}}_\ell$ with $\dim \mathrm{Fil}^0 = 5$ and $\dim \mathrm{Fil}^1 = 1$. Hence $\mathrm{Fil}^0 \cap H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))[1/\ell]^{\varphi=1}$ is a $\mathbb{Q}_\ell$-vector space of dimension at most 5. By Grothendieck–Messing theory, both $\mathcal{O}_F$ and Λ_v lie in Fil^0 and hence $m' + 2 \leq 5$. If $m' = 3$, then $\mathrm{Span}\{\mathcal{O}_F, \Lambda_v\} = \mathrm{Fil}^0$. By Mazur's weak admissibility theorem, since both $\mathcal{O}_F$ and Λ_v lie in $H^2_{\mathrm{cris}}(\mathcal{B}_{v,1}/W(\bar{\mathbb{F}}_\ell))[1/\ell]^{\varphi=1}$, it follows that $\mathrm{Span}\{\mathcal{O}_F, \Lambda_v\}$ has trivial filtration. This contradicts $0 \neq \mathrm{Fil}^1 \subset \mathrm{Fil}^0$, and we conclude that $m' \leq 2$. □

Theorem 4.3.3. *Let $M(N, n, \epsilon)$ denote the number of primes $p \in [N^{1/2}, N]$ such that $\#\{[\mathcal{B}_{v,n}] \in T_{\mathfrak{p}}([\mathcal{A}_{v,n}]) \mid [\mathcal{B}_{v,n}] \in \bigcup_{r\in\mathbb{I}} \mathcal{T}(r)\} \geq \epsilon p$. Then $M(N, \lceil 3e_v \log\log N\rceil, \epsilon) = o_\epsilon(N/(\log N))$.*

Proof. The number of primes in the interval $[N^{1/2}, N/\log(N)]$ is $o(N/\log(N))$, so we will restrict ourselves to primes $p \in [N/\log(N), N]$. For each prime p, each $[\mathcal{B}_{v,n}] \in T_{\mathfrak{p}}([\mathcal{A}_{v,n}])$ which lies in $\cup_{r\in\mathbb{I}}\mathcal{T}(r)$ induces a special endomorphism of $\mathcal{A}_{v,n}$ whose norm is $pr\,\mathrm{Nm}\,\mathfrak{a}/D$. For all $p \in [N/\log N, N]$, the quantity $pr\,\mathrm{Nm}\,\mathfrak{a}/D = O(N)$. Notice that distinct $[\mathcal{B}_{v,n}] \in T_{\mathfrak{p}}([\mathcal{A}_{v,n}])$ which lie in $\cup_{r\in\mathbb{I}}\mathcal{T}(r)$ induce distinct special endomorphisms of $\mathcal{A}_{v,n}$. Therefore, $\mathcal{A}_{v,n}$ has at least $M(N, n, \epsilon)\epsilon N/\log N$ special endomorphisms with norm bounded by $O(N)$.

Applying the crudest bounds that Theorem 4.1.1, Lemma 4.2.1 and Lemma 4.3.2 yield, the number of special endomorphisms of $\mathcal{A}_{v,n}$ with norm bounded by $O(N)$ is $O\left(\frac{N^2}{\ell^{2n/e_v}} + N^{3/2}\right)$. Therefore $M(N, n, \epsilon) = O_\epsilon\left(\frac{N\log N}{\ell^{2n/e_v}} + N^{1/2}\log N\right)$. Substituting $n = \lceil 3e_v \log\log N\rceil$ yields $M(N, n, \epsilon) = o_\epsilon(N/(\log N))$ as required. □

The following theorem shows that one can choose a sequence of p such that the largest v-adic intersection multiplicity of a point in $T_{\mathfrak{p}}([\mathcal{A}])$ with $\bigcup_{r\in\mathbb{I}}\mathcal{T}(r)$ is $O(\log p)$.

Theorem 4.3.4. *Set $n = \lceil \frac{e_v \log N}{\log \ell}\rceil$. Then the number of primes $p \in [N^{1/2}, N]$ for which there exists $[\mathcal{B}] \in T_{\mathfrak{p}}([\mathcal{A}])$ and $r \in \mathbb{I}$ with $[\mathcal{B}_{v,n}] \in \mathcal{T}(r)$ is $o(N/(\log N))$.*

Proof. Let $p \in [N^{1/2}, N]$ such that there exists $[\mathcal{B}] \in T_{\mathfrak{p}}([\mathcal{A}])$ and $r \in \mathbb{I}$ as in the statement. Then $\mathcal{A}_{v,n}$ has a special endomorphism, say s_p, of norm $pr\,\mathrm{Nm}\,\mathfrak{a}/D$. Clearly, $s_p \neq s_{p'}$ where $p' \neq p$ also satisfies the conditions in the statement. Therefore, each such p induces a distinct special endomorphism of $\mathcal{A}_{v,n}$ having norm $O(N)$ and it suffices to bound the number of special endomorphisms of norm $\leq O(N)$.

Let Λ_v denote the module of special endomorphisms of $\mathcal{A}_{\mathcal{O}_{K_v}}[\ell^\infty]$. If Λ_v has $\mathbb{Z}_\ell$-rank ≤ 1, then Corollary 4.2.3 yields the desired result. Therefore, we assume that the rank is at least 2. By Lemma 4.3.2, the rank of Λ_v is at most 2, so we assume that the rank equals 2. Let $n = n_0' + ke_v$, where $n_0' - n_0 < e_v$ and n_0 is as in Theorem 4.1.1. We first deal with the case when $\mu_1(M_{v,n})\mu_2(M_{v,n}) \geq (\log N)^2$. By Lemma 4.2.1 $\#\{s \in M_{v,n} : Q(s) \leq O(N)\} = O(\frac{N^2}{\ell^{2n/e_v}} + \frac{N^{3/2}}{\ell^{n/e_v}} + \frac{N}{\mu_1(M_{v,n})\mu_2(M_{v,n})} + N^{1/2})$.[17] This quantity is $o(N/\log N)$ and so the result follows in this case.

We may therefore assume that $\mu_1(M_{v,n})\mu_2(M_{v,n}) \leq (\log N)^2$. Let $v_1, v_2 \in M_{v,n}$ denote two linearly independent vectors with length $\mu_1(M_{v,n}), \mu_2(M_{v,n})$ respectively and let L_n denote the rank-2 sublattice spanned by the two vectors, and let L_n' denote its saturation in $M_{v,n}$. We first claim that the index of L_n in L_n' is bounded by an absolute constant. To prove this claim, we notice that the lattices L_n' and L_n have the same successive minima: $\mu_i(M_{v,n}) \leq \mu_i(L_n') \leq \mu_i(L_n) = \mu_i(M_{v,n})$ for $i = 1, 2$, with the last equality following by construction. By [EK95, Equations 5,6], $\mu_1(M_{v,n})\mu_2(M_{v,n}) \ll d(L_n') \ll \mu_1(M_{v,n})\mu_2(M_{v,n})$, which yields that $d(L_n) \ll d(L_n') \leq d(L_n)$.

We claim that any vector $v \in M_{v,n}$ such that $Q(v) \leq O(N)$ actually lies in L_n'. Indeed, if this were not the case, then the vectors v_1, v_2, v are linearly independent, and the length of v is bounded above by $O(N^{1/2})$. Therefore, it follows that $\mu_3(M_{v,n}) \leq O(N^{1/2})$. This is a contradiction. Indeed, by Lemma 4.2.2 and the assumption that $\mu_1(M_{v,n})\mu_2(M_{v,n}) \leq \log(N)^2$, we have that $\mu_3(M_{v,n}) \gg \ell^{n/e_v}/(\log N)^2 \geq N/(\log N)^2 > O(N^{1/2})$. Therefore, we may assume that every vector $v \in M_{v,n}$ with $Q(v) \leq O(N)$ lies in L_n'.

Finally, we claim that $d(L_n') \to \infty$ as $n \to \infty$. As $d(L_n) \ll d(L_n') \leq d(L_n)$, it suffices to prove that $d(L_n) \to \infty$. Clearly, $\mu_1(M_{v,n}) \leq \mu_1(L_n) \ll d(L_n)$. Given any $C > 0$, there are only finitely many vectors $v \in M_{v,1}$ with length bounded by C. As the intersection $\bigcap_{i=1}^\infty M_{v,i} = \{0\}$ (this is because $\mathcal{A}$ has no special

[17] Note that k differs from n/e_v by a quantity bounded independent of n, so $1/\ell^k = O(1/\ell^{\frac{n}{e_v}})$.

endomorphisms generically), it follows that for i large enough, $M_{v,i}$ contains no vectors with length bounded by C, and hence for i large enough, $\mu_1(M_{v,i}) > C$. Hence, $\mu_1(M_{v,n}) \to \infty$, and therefore $d(L_n) \to \infty$. Now, the same argument used to prove Claim 3.1.9 applies to prove that the proportion of primes p, such that there exists a $\mathcal{B} \in T_{\mathfrak{p}}(\mathcal{A})$ modulo v^n, goes to zero. □

5. Proof of the main theorem

The goal of this section is to deduce our main theorem from the results in §§3-4 which provide upper bounds of the local intersection numbers. Recall that A is an abelian surface over K with an $\mathfrak{a}$-polarization and $\mathcal{O}_F \subseteq \mathrm{End}(A)$. As in 2.1, we will assume the existence of a semi-abelian scheme $\mathcal{A}$ over $\mathcal{O}_K$ with semistable reduction everywhere, whose generic fiber is A. Recall that p denotes a prime which is totally split in the narrow Hilbert class field of F. To prepare for our proof, we first use Borcherds' theory to choose a suitable Hirzebruch–Zagier divisor in the Hilbert modular surface and then compute the asymptotic of Faltings heights on the Hecke orbits.

5.1. **Borcherds' theory and the Faltings height.** We devote this subsection to applying arithmetic Borcherds' theory to choose a rational section of certain tensor powers of the Hodge line bundle. We then interpret the Faltings height of an abelian surface as a certain Arakelov intersection number.

We use $\mathcal{A}^{\text{univ-sa}}$ to denote the universal family of semi-abelian schemes over $\overline{\mathcal{H}}^{\text{tor}}$ (with suitable level structure). In [BBGK07, §6], the authors explain a way to define the arithmetic intersection independently of the choice of a level structure. We recall their definition in 5.1.3. Let $e : \overline{\mathcal{H}}^{\text{tor}} \to \mathcal{A}^{\text{univ-sa}}$ be the identity section and let $\omega = \det(e^*\Omega^1_{\mathcal{A}^{\text{univ-sa}}/\overline{\mathcal{H}}^{\text{tor}}})$ over $\overline{\mathcal{H}}^{\text{tor}}$ be the Hodge line bundle. We endow ω with a Hermitian metric $||\cdot||_F$ (only on $\mathcal{H}(\mathbb{C})$) as in [Fal86, sec. 3] and denote by $\overline{\omega}$ the Hermitian line bundle with log singularity along the boundary. By definition, we have

$$h_F(A) = ht_{\overline{\omega}}([\mathcal{A}]),$$

where h_F denotes the stable Faltings height and ht is the height function of subvarieties of an arithmetic variety with respect to certain arithmetic cycles (see, for example, [BBGK07, §1.5, eqn. (1.17)] and we normalize h_F and $ht_{\overline{\omega}}$ to be independent on the choice of K; more specifically, $||\ell||_v = \ell^{-\frac{[K_v:\mathbb{Q}_\ell]}{[K:\mathbb{Q}]}}$).

It is well known that the space of global sections of the line bundle $\omega^{\otimes k}$ over $\overline{\mathcal{H}}^{\text{tor}}_{\mathbb{C}}$ (resp. $\overline{\mathcal{H}}^{\text{tor}}_{\overline{\mathbb{Q}}}$) is the space of Hilbert modular forms of parallel weight k over $\mathbb{C}$ (resp. $\overline{\mathbb{Q}}$); see, for example, [FC90, Chp. V.1] and [Cha90, sec. 4]. Up to a constant, the Hermitian metric[18] $||\cdot||_F$ on $\omega^{\otimes k}$ is defined by $||f(z)||_{\text{Pet}} = |f(z_1, z_2)(\Im z_1)^{k/2}(\Im z_2)^{k/2}|$, where f is a Hilbert modular form of parallel weight k and $z = (z_1, z_2) \in \mathbb{H}^2$: indeed, this follows from the $G(\mathbb{R})$-invariance of both metrics.

Lemma 5.1.1. *There exist a positive integer k and a meromorphic Hilbert modular form Ψ over $\overline{\mathbb{Q}}$ of parallel weight k such that the divisor* $\mathrm{Div}(\Psi)$ *defined by Ψ on $\overline{\mathcal{H}}^{\text{tor}}_{\overline{\mathbb{Q}}}$ is given by $\sum_{r\in\mathbb{I}} c_r T(r)$, where $c_r \in \mathbb{Z}$ and $\mathbb{I}$ is a finite subset of*

$$\mathbb{J} = \{qD \mid q \text{ is a rational prime inert in } F\}.$$

In particular, $\mathrm{Div}(\Psi)$ *is a weighted sum of compact Shimura curves.*

We use Borcherds' theory [Bor98] to construct meromorphic Hilbert modular forms on $\mathcal{H}_{\mathbb{C}}$ from certain weakly holomorphic modular forms on modular curves. The Hilbert modular form constructed in such a way (sometimes called a Borcherds lift) has the following properties: (1) When viewed as a rational section of certain tensor power of the Hodge line bundle, the divisor defined by the Hilbert modular form is a linear combination of $T(r)$ determined by the principal part of the Fourier expansion of the given weakly holomorphic modular form; (2) Divide $\mathbb{H}^2$ into Weyl chambers and fix any cusp. The Hilbert modular form can be explicitly written down as an infinite product (sometimes called a Borcherds product) near the cusp. Borcherds also showed that the existence of such lifts of a weakly holomorphic modular form can be verified by certain explicit conditions on the Fourier coefficients of its principal part. The Fourier expansions of Borcherds lifts have also been studied by many people, leading to an arithmetic theory of these lifts. See [BBGK07, §4] for a summary of relevant results when the discriminant of F is a prime.

[18]We also use $||\cdot||_F$ to denote the metric on $\overline{\omega}^{\otimes k}$ given by the tensor product of the Hermitian metric $||\cdot||_F$ on $\overline{\omega}$.

Proof of Lemma 5.1.1. By [Bru16, Thm. 1.1],[19] in which we take the infinite admissible set to be $\mathbb{J}$, there exists a Borcherds product Ψ' of non-zero weight k whose divisor is supported on $\cup_{r\in\mathbb{J}}T(r)$. In other words, Ψ' is a Hilbert modular form of parallel weight k over $\mathbb{C}$. We may assume $k>0$, since otherwise we just take Ψ'^{-1}. By [Bru16, Prop. 3.1], the weakly holomorphic modular f whose Borcherds lift is Ψ has integral Fourier coefficients. [Hör14, 3.2.14] shows that, after multiplying by a suitable scalar, the Borcherds lift of a modular form with Fourier coefficients in $\mathbb{Q}$ is defined over $\overline{\mathbb{Q}}$. In particular, if we take Ψ to be Ψ' multiplied by a suitable scalar, then Ψ is a rational section of $\omega^{\otimes k}$ over $\overline{\mathcal{H}}^{\mathrm{tor}}_{\overline{\mathbb{Q}}}$. For $r\in\mathbb{J}$, the divisor $T(r)$ is compact by Corollary 2.1.3. At each cusp, we use the Borcherds product to study Ψ. Since $T(r)$ is compact, the Weyl chamber is the whole $\mathbb{H}^2$ and the Borcherds product expression of Ψ does not vanish at the cusps. □

5.1.2. We view Ψ as a rational section of $\omega^{\otimes k}$ over $\overline{\mathcal{H}}^{\mathrm{tor}}_{\mathcal{O}_{K'}}$, where K' is a large enough number field such that Ψ is defined. Hence $\mathrm{Div}(\Psi)=\sum_r c_r\mathcal{T}(r)+\sum_p\mathcal{E}_p$, where the second sum is over finitely many p and $\mathcal{E}_p$ is a finite (weighted) sum of irreducible components of $\overline{\mathcal{H}}^{\mathrm{tor}}_{\overline{\mathbb{F}}_p}$.

5.1.3. Given an arithmetic divisor $\mathcal{D}$ and a horizontal 1-cycle $\mathcal{Z}$ intersecting properly on $\overline{\mathcal{H}}^{\mathrm{tor}}_{\mathcal{O}_{K'}}$, one defines the (arithmetic) intersection number as follows: (see, for example, [BGKK07, Thm. 1.33][20] for regular schemes and [Yan10, eqn. (2.1)] for regular Deligne–Mumford stacks.)

$$\mathcal{D}.\mathcal{Z}=\sum_v\sum_{x\in(\mathcal{Z}\cap\mathcal{D})(\overline{\mathbb{F}}_v)}\frac{\log(\#\tilde{\mathcal{O}}_{\mathcal{Z}\cap\mathcal{D},x})}{\#\operatorname{Aut}(x)}=\sum_v\sum_{x\in(\mathcal{Z}\cap\mathcal{D})(\overline{\mathbb{F}}_v)}\frac{\mathrm{Length}(\tilde{\mathcal{O}}_{\mathcal{Z}\cap\mathcal{D},x})\log(\#k(x))}{\#\operatorname{Aut}(x)},$$

where v ranges over the finite places of K', the intersection $\mathcal{Z}\cap\mathcal{D}=\mathcal{Z}\times_{\overline{\mathcal{H}}^{\mathrm{tor}}}\mathcal{D}$ is a Deligne–Mumford stack of dimension 0, the ring $\widetilde{\mathcal{O}}_{\mathcal{Z}\cap\mathcal{D},x}$ is the strictly Henselian local ring of $\mathcal{Z}\cap\mathcal{D}$ at x, and $k(x)$ is the residue field of $\widetilde{\mathcal{O}}_{\mathcal{Z}\cap\mathcal{D},x}$. By definition, $\dfrac{\mathcal{D}.\mathcal{Z}}{[K':\mathbb{Q}]}$ is independent of the choice of K'.

In [BBGK07, sec. 6.3], they define the arithmetic intersection number on $\overline{\mathcal{H}}^{\mathrm{tor}}$ as the arithmetic intersection number of the pull back of arithmetic cycles to $\overline{\mathcal{H}}^{\mathrm{tor}}(N)$, the Hilbert modular surface with full level N-structure with $N\geq 3$, divided by the degree of the map $\overline{\mathcal{H}}^{\mathrm{tor}}(N)\to\overline{\mathcal{H}}^{\mathrm{tor}}$. This is the idea behind the above formula.

Remark 5.1.4. For $\mathcal{D}$ and $\mathcal{Z}$ as above, let n denote the largest integer such that $\mathcal{Z}$ is contained in $\mathcal{D}$ modulo v^n (here, we consider $\mathcal{Z}$ and $\mathcal{D}$ as subschemes of the coarse Hilbert modular surface). In our applications, we will only consider the intersection at finitely many places. Therefore, we may pass to a suitable level structure étale at these finitely many places so that $\mathcal{T}(r)$ are regular ([Car86]). Then the length at v referred to in 5.1.3 differs from n by an absolutely bounded factor. As we are only concerned with bounds, we will in the sequel restrict ourselves with controlling the growth of n.

Lemma 5.1.5. *Assume that* $\mathrm{End}(A_{\overline{K}})=\mathcal{O}_F$. *Following the notation as in Lemma 5.1.1, there exists a constant* C_1 *independent of* A *such that*

$$\left|h_F(A)-\frac{1}{k[K:\mathbb{Q}]}[\mathcal{A}].\sum_{r\in\mathbb{I}}c_r\mathcal{T}(r)+\frac{1}{k[K:\mathbb{Q}]}\sum_{\sigma:K\hookrightarrow\mathbb{C}}\frac{1}{\#\operatorname{Aut}(A_{\overline{K}})}\log||\Psi(\sigma([A]))||_{\mathrm{Pet}}\right|<C_1.$$

Proof. By Lemma 2.1.6, if $[A]$ lies on $T(r)$, then $\mathrm{End}(A_{\overline{K}})$ is strictly larger than $\mathcal{O}_F$. Hence our assumption implies that $[A]$ does not lie on any $T(r)$. It follows that the 1-cycle $[\mathcal{A}]$ intersects $\sum_{r\in\mathbb{I}}c_rT(r)$ properly. Set $\mathcal{E}(A)=\frac{1}{[K:\mathbb{Q}]}[\mathcal{A}].(\sum_p\mathcal{E}_p)$, where $\sum_p\mathcal{E}_p$ was defined in 5.1.2. By definition, $\mathcal{E}(A)$ is bounded by an absolute constant independent of A. Moreover,

$$h_F(A)=ht_{\overline{\omega}}([\mathcal{A}])=\frac{1}{k[K:\mathbb{Q}]}[\mathcal{A}].\sum_{r\in\mathbb{I}}c_r\mathcal{T}(r)-\frac{1}{k[K:\mathbb{Q}]}\sum_{\sigma:K\hookrightarrow\mathbb{C}}\frac{1}{\#\operatorname{Aut}(A_{\overline{K}})}\log||\Psi(\sigma([A]))||_F+\frac{1}{k}\mathcal{E}(A),$$

(see, for example, [Yan10, eqn. (2.3)]). The lemma then follows from the fact that $||\cdot||_F$ and $||\cdot||_{\mathrm{Pet}}$ differ by an absolute constant independent of A. □

[19] [Bru16, Thm. 1.1] is a generalization of [BBGK07, Lem. 4.11]. The proof of this lemma, which only deals with the case when D is a prime, contains the main idea of the proof for the general case.

[20] The higher tor group vanishes since we work with a Cartier divisor which intersects the 1-cycle properly.

We end this subsection with a formula for the average Faltings height of abelian surfaces corresponding to points in $T_{\mathfrak{p}}[A]$ when A has good reduction at all the primes of K above p. The idea of proof builds on Autissier's idea in [Aut05]. From now on, we say that A has *good reduction at p* (resp. *ordinary reduction at p*) if A has good reduction (resp. ordinary reduction) at all the primes of K above p.

Proposition 5.1.6. *Let p be a prime as in 2.2.1. If A has good reduction at p, then*

$$\sum_{[B]\in T_{\mathfrak{p}}[A]} h_F(B) = (p+1)h_F(A) + \frac{p-1}{2}\log p.$$

Proof. The proof consists two parts. We first show that $\sum_{[B]\in T_{\mathfrak{p}}[A]} h_F(B) - (p+1)h_F(A)$ is independent of A. Then we compute this quantity in a particular case.

(1) Let $\mathcal{H}(\mathfrak{p})$ denote the Hilbert modular surface over $\mathbb{Z}$ with $\Gamma_0(\mathfrak{p})$ level structure. The stack $\mathcal{H}(\mathfrak{p})$ parametrizes degree p isogenies $\phi : A_1 \to A_2$ between abelian varieties with $\mathcal{O}_F$-multiplication such that $\ker(\phi) \subset A_1[\mathfrak{p}]$ (see for example [Pap95, sec. 2.2]). Let $\pi_i : \mathcal{H}(\mathfrak{p})_{\mathbb{Z}_{(p)}} \to \mathcal{H}_{\mathbb{Z}_{(p)}}$ for $i = 1, 2$ be the forgetful map that sends ϕ to A_i. We claim that each π_i is finite flat.

We first show that π_i is quasi-finite. Let v be any finite place of K over p. The group scheme $\mathcal{A}[\mathfrak{p}^\infty]$ of $\mathfrak{p}$-power torsions of $\mathcal{A}$ is a p-divisible group of height 2, whose mod v reduction has dimension 1. There are two cases: the mod-v reduction of $\mathcal{A}[\mathfrak{p}^\infty]$ is either ordinary, or supersingular. If ordinary, [FC90, §VII.4] shows that there are only finitely many degree p subgroups of the mod-v reduction of $\mathcal{A}[\mathfrak{p}]$. Therefore, the modular interpretation of $\mathcal{H}(\mathfrak{p})$ shows that the map π_1 is quasi-finite. Now we assume that the mod-v reduction of $\mathcal{A}[\mathfrak{p}^\infty]$ is supersingular. Since the number of degree p subgroups of the reduction of $\mathcal{A}[\mathfrak{p}]$ only depends on the isomorphism class of the p-divisible group, we may assume that $\mathcal{A}$ has supersingular reduction at v. Since p is split in F, the supersingular locus of $\mathcal{H}_{\mathbb{F}_v}$ is 0-dimensional and hence there are only finitely many mod-v points on $\mathcal{H}$ corresponding to abelian surfaces isogenous to $\mathcal{A}$ mod v. In particular, π_1 is quasi-finite. The same argument applies to π_2 when we study the kernel of the Rosati involution of ϕ.

By [Pap95, 2.1.3, Cor. 2.2.3], the stack $\mathcal{H}(\mathfrak{p})$ is Cohen–Macaulay and $\mathcal{H}$ is regular. Since all the fibers of π_i are 0-dimensional, it follows by [EGAIV, II.6.1.5] that π_i is flat. On the other hand, each π_i is proper by [Pap95, the discussion after Def. 2.2.1]. Therefore, each π_i is finite flat.

By the argument in [Aut05, Theorem 5.1], one observes that the independence of the quantity $\sum_{[B]\in T_{\mathfrak{p}}[A]} h_F(B) - (p+1)h_F(A)$ on A is a formal consequence of π_i being finite-flat, of $\mathcal{H}$ being normal, and of $\mathcal{H}_{\mathbb{F}_v}$ being irreducible for every v above p.

(2) We now compute $\sum_{[B]\in T_{\mathfrak{p}}[A]} h_F(B) - (p+1)h_F(A)$ when A has ordinary reduction at p.[21] Such an A always exists: indeed, for a CM field K_2 containing F such that K_2 is Galois over $\mathbb{Q}$ of degree 4 and p splits completely in K_2, there exist abelian surfaces with CM by K_2 that correspond to points on $\mathcal{H}$ and these abelian surfaces are ordinary at p.

Now assume that A has good ordinary reduction at p and we prove the result for such A. We first enlarge K so that all $[B] \in T_{\mathfrak{p}}[A]$ are defined over K. Since $\mathcal{A}[\mathfrak{p}^\infty]$ over $\mathcal{O}_{K_v}$ is 1-dimensional, it only contains a unique degree p subgroup which is multiplicative (equivalently, in the ordinary case, not étale) for any v above p. Then by Lemma 2.2.2, there exists only one element in $T_{\mathfrak{p}}[\mathcal{A}]$ that corresponds to an isogeny with multiplicative kernel. Now we apply [Fal86, Lem. 5]. By the standard calculation on Ω^1 of finite flat groups of degree p, at each v there are p out of $p+1$ elements in $T_{\mathfrak{p}}[A]$ such that the term $\log(\#e^*(\Omega^1_{\ker\phi/\mathcal{O}_{K_v}}))$ in Faltings' formula is 0, and one element such that $\log(\#e^*(\Omega^1_{\ker\phi/\mathcal{O}_{K_v}})) = [K_v : \mathbb{Q}_p]\log p$. We obtain the desired formula by summing up all the local contributions. □

5.2. **Proof of Theorem 1.**

5.2.1. We first sketch the proof of Theorem 1. First, we use Lemma 5.1.1 to choose a good Hirzebruch–Zagier divisor $\sum_{r\in\mathbb{I}} c_r\mathcal{T}(r) = \mathrm{Div}(\Psi)$. By Proposition 5.1.6, we have $\sum_{[B]\in T_{\mathfrak{p}}([A])} h_F(B) \gg p\log p$. On the other hand, the local results in §§3-4 show that for most of primes p, each local term in Lemma 5.1.5 is $o(p\log p)$.

[21] One may also choose any CM abelian surface on $\mathcal{H}$ and apply the formula for the Faltings height in [Moc17] to compute this difference.

By local term, we mean either $-\sum_{[B]\in T_{\mathfrak{p}}[A]} \log ||\Psi(\sigma([B]))||_{\mathrm{Pet}}$ for all $\sigma : K \hookrightarrow \mathbb{C}$ or the v-adic intersection number $(\sum_{[\mathcal{B}]\in T_{\mathfrak{p}}[\mathcal{A}]}[\mathcal{B}], \sum_{r\in\mathbb{I}} c_r\mathcal{T}(r))_v$ for finite places v. This implies that $T_{\mathfrak{p}}([\mathcal{A}])$ intersects $\sum_{r\in\mathbb{I}} c_r\mathcal{T}(r)$ at infinitely many places as $p \to \infty$ and then Theorem 1 follows from Corollary 2.1.7.

5.2.2. If $\mathcal{O}_F \subsetneq \mathrm{End}(A_{\overline{K}})$, then by the classification of the endomorphism ring of absolutely simple abelian surfaces over a characteristic zero field, $\mathrm{End}(A_{\overline{K}}) \otimes \mathbb{Q}$ is either an indefinite quaternion algebra over $\mathbb{Q}$ or a degree 4 CM field. In the first case, $\mathcal{A}_{\overline{\mathbb{F}}_v}$ is not simple if the quaternion algebra splits at $\mathrm{char}\,\mathbb{F}_v$. In the second case, there exists a positive density set of primes ℓ so that $\mathcal{A}_{\mathbb{F}_v}$ is supersingular for all $v|\ell$ and hence not geometrically simple. Therefore, to prove Theorem 1, we assume that $\mathrm{End}(A_{\overline{\mathbb{Q}}}) = \mathcal{O}_F$ from now on and hence for any $[B] \in T_{\mathfrak{p}}([A])$, we also have $\mathrm{End}(B_{\overline{K}}) = \mathcal{O}_F$. Therefore, all $T_{\mathfrak{p}}([\mathcal{A}])$ intersect Hirzebruch–Zagier divisors properly.

Proof of Theorem 1. We now show that there are infinitely many primes v of K such that $\mathcal{A}_{\overline{\mathbb{F}}_v}$ is not simple. Assume, for the sake of contradiction, that there is a finite set of places Σ of K such that $\mathcal{A}$ has geometrically simple, or bad reduction modulo v for $v \notin \Sigma$. Corollary 2.1.3 and Lemma 5.1.1 give a meromorphic Hilbert modular form Ψ such that $\mathrm{Div}(\Psi)$ is a compact special divisor $\sum_{r\in\mathbb{I}} c_r\mathcal{T}(r)$ with $D|r$ for all $r \in \mathbb{I}$. The intersection $(T_{\mathfrak{p}}([\mathcal{A}]), \sum_{r\in\mathbb{I}} c_r\mathcal{T}(r))$ has a nonzero v-adic term only when $v \in \Sigma$ by Corollary 2.1.7. Throughout the proof, p will denote a prime which is totally split in the narrow Hilbert class field of F and $v \nmid p$ for all $v \in \Sigma$. We now explain how to choose an increasing sequence of primes p such that we can bound the local terms as described in 5.2.1 by $\epsilon p \log p$ for arbitrary $\epsilon > 0$.

By Theorems 4.3.3 and 4.3.4 and Remark 5.1.4, outside a density-zero set of primes p, the v-adic intersection

$$\Big(\sum_{[\mathcal{B}]\in T_{\mathfrak{p}}[\mathcal{A}]}[\mathcal{B}], \sum_{r\in\mathbb{I}} c_r\mathcal{T}(r)\Big)_v \le C_1\bigg(\sum_{r\in\mathbb{I}} |c_r| \cdot \Big((p+1)(3e_v\log(2\log p)) + \epsilon p(2e_v\log p + 1)\Big)\bigg),$$

where C_1 is the absolute constant mentioned in Remark 5.1.4. Notice that $2\log p \ge \log N$.

Let C_2 be the density of primes splitting completely in the narrow Hilbert class field of F. By taking $\epsilon_1 = \epsilon$, $\epsilon_2 = \frac{C_2}{4[K:\mathbb{Q}]}$ in Theorem 3.2.1, we have that, for $N \gg 0$ and for $p \in [N^{1/2}, N]$ in a set of density at least $3C_2/4$, for all $\sigma : K \hookrightarrow \mathbb{C}$, the archimedean term

$$-\sum_{[B]\in T_{\mathfrak{p}}[A]} \log ||\Psi(\sigma([B]))||_{\mathrm{Pet}} < \epsilon p \log p.$$

We have shown that for $N \gg 1$, there exists a positive density set of primes $p \in [N^{1/2}, N]$ such that all the local terms are $o(p\log p)$. On the other hand, by Proposition 5.1.6, $\sum_{[B]\in T_{\mathfrak{p}}([A])} h_F(B)$ has order of magnitude $p\log p$. We then obtain the desired contradiction by applying Lemma 5.1.5 to all $[B] \in T_{\mathfrak{p}}([A])$. □

Remark 5.2.3. A slight modification of the proof shows that for any indefinite non-split quaternion algebra D over $\mathbb{Q}$ such that $F \subset D$, there exists infinitely many places v of K such that $D \subset \mathrm{End}^0(A_{\overline{\mathbb{F}}_v})$. Indeed, we may take $\mathbb{J} = \{r \in \mathbb{Z} \mid T(r) \text{ is a Shimura curve with associated quaternion algebra to be } D\}$ and the rest of the proof follows.

References

[AH17] Jeffrey D. Achter and Everett W. Howe, *Split abelian surfaces over finite fields and reductions of genus-2 curves*, Algebra Number Theory **11** (2017), no. 1, 39–76.

[Ach12] Jeffrey D. Achter, *Explicit bounds for split reductions of simple abelian varieties*, J. Théor. Nombres Bordeaux **24** (2012), no. 1, 41–55 (English, with English and French summaries).

[Aut05] Pascal Autissier, *Hauteur moyenne de variétés abéliennes isogènes*, Manuscripta Math. **117** (2005), no. 1, 85–92 (French, with English and French summaries).

[Bru16] Jan Hendrik Bruinier, *Borcherds products with prescribed divisor* (2016). Available on arXiv: 1607.08713.

[Bor98] Richard E. Borcherds, *Automorphic forms with singularities on Grassmannians*, Invent. Math. **132** (1998), no. 3, 491–562.

[BBGK07] Jan H. Bruinier, José I. Burgos Gil, and Ulf Kühn, *Borcherds products and arithmetic intersection theory on Hilbert modular surfaces*, Duke Math. J. **139** (2007), no. 1, 1–88.

[BGKK07] J. I. Burgos Gil, J. Kramer, and U. Kühn, *Cohomological arithmetic Chow rings*, J. Inst. Math. Jussieu **6** (2007), no. 1, 1–172.

[Car86] Henri Carayol, *Sur la mauvaise réduction des courbes de Shimura*, Compositio Math. **59** (1986), no. 2, 151–230 (French).

[Cha90] C.-L. Chai, *Arithmetic minimal compactification of the Hilbert-Blumenthal moduli spaces*, Ann. of Math. (2) **131** (1990), no. 3, 541–554.

[Cha14] Francois Charles, *Exceptional isogenies between reductions of pairs of elliptic curves* (2014). Available at arXiv:1411.2914.

[Cha97] Nick Chavdarov, *The generic irreducibility of the numerator of the zeta function in a family of curves with large monodromy*, Duke Math. J. **87** (1997), no. 1, 151–180.

[Chi92] Wên Chên Chi, *l-adic and λ-adic representations associated to abelian varieties defined over number fields*, Amer. J. Math. **114** (1992), no. 2, 315–353.

[COU01] Laurent Clozel, Hee Oh, and Emmanuel Ullmo, *Hecke operators and equidistribution of Hecke points*, Invent. Math. **144** (2001), no. 2, 327–351.

[Del79] Pierre Deligne, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 247–289 (French).

[Elk87] Noam D. Elkies, *The existence of infinitely many supersingular primes for every elliptic curve over* $\mathbf{Q}$, Invent. Math. **89** (1987), no. 3, 561–567, DOI 10.1007/BF01388985.

[Elk89] ———, *Supersingular primes for elliptic curves over real number fields*, Compositio Math. **72** (1989), no. 2, 165–172.

[EK95] Alex Eskin and Yonatan R. Katznelson, *Singular symmetric matrices*, Duke Math. J. **79** (1995), no. 2, 515–547, DOI 10.1215/S0012-7094-95-07913-7.

[Fal86] Gerd Faltings, *Finiteness theorems for abelian varieties over number fields*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 9–27. Translated from the German original [Invent. Math. **73** (1983), no. 3, 349–366; ibid. **75** (1984), no. 2, 381; MR 85g:11026ab] by Edward Shipz.

[FC90] Gerd Faltings and Ching-Li Chai, *Degeneration of abelian varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 22, Springer-Verlag, Berlin, 1990. With an appendix by David Mumford.

[FKRS12] Francesc Fité, Kiran S. Kedlaya, Víctor Rotger, and Andrew V. Sutherland, *Sato-Tate distributions and Galois endomorphism modules in genus 2*, Compos. Math. **148** (2012), no. 5, 1390–1442, DOI 10.1112/S0010437X12000279.

[Gor02] Eyal Z. Goren, *Lectures on Hilbert modular varieties and modular forms*, CRM Monograph Series, vol. 14, American Mathematical Society, Providence, RI, 2002. With the assistance of Marc-Hubert Nicole.

[EGAIV] A. Grothendieck, *Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. I*, Inst. Hautes Études Sci. Publ. Math. **20** (1964), 259 (French).

[HY12] Benjamin Howard and Tonghai Yang, *Intersections of Hirzebruch-Zagier divisors and CM cycles*, Lecture Notes in Mathematics, vol. 2041, Springer, Heidelberg, 2012.

[HZ76] F. Hirzebruch and D. Zagier, *Intersection numbers of curves on Hilbert modular surfaces and modular forms of Nebentypus*, Invent. Math. **36** (1976), 57–113.

[Hör14] Fritz Hörmann, *The geometric and arithmetic volume of Shimura varieties of orthogonal type*, CRM Monograph Series, vol. 35, American Mathematical Society, Providence, RI, 2014.

[KS99] Nicholas M. Katz and Peter Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, vol. 45, American Mathematical Society, Providence, RI, 1999.

[Ked15] Kiran S. Kedlaya, *Sato-Tate groups of genus 2 curves*, Advances on superelliptic curves and their applications, NATO Sci. Peace Secur. Ser. D Inf. Commun. Secur., vol. 41, IOS, Amsterdam, 2015, pp. 117–136.

[Kis10] Mark Kisin, *Integral models for Shimura varieties of abelian type*, J. Amer. Math. Soc. **23** (2010), no. 4, 967–1012.

[KR99] Stephen S. Kudla and Michael Rapoport, *Arithmetic Hirzebruch-Zagier cycles*, J. Reine Angew. Math. **515** (1999), 155–244.

[LT76] Serge Lang and Hale Trotter, *Frobenius distributions in* GL_2*-extensions*, Lecture Notes in Mathematics, Vol. 504, Springer-Verlag, Berlin-New York, 1976. Distribution of Frobenius automorphisms in GL_2-extensions of the rational numbers.

[Mes72] William Messing, *The crystals associated to Barsotti-Tate groups: with applications to abelian schemes*, Lecture Notes in Mathematics, Vol. 264, Springer-Verlag, Berlin-New York, 1972.

[MP08] V. Kumar Murty and Vijay M. Patankar, *Splitting of abelian varieties*, Int. Math. Res. Not. IMRN **12** (2008), Art. ID rnn033, 27, DOI 10.1093/imrn/rnn033.

[Moc17] Lucia Mocz, *A new Northcott property for Faltings height* (2017). preprint.

[Ogu82] Arthur Ogus, *Hodge cycles and crystalline cohomology*, Hodge cycles, motives, and Shimura varieties, Lecture Notes in Mathematics, vol. 900, Springer-Verlag, Berlin-New York, 1982.

[Pap95] Georgios Pappas, *Arithmetic models for Hilbert modular varieties*, Compositio Math. **98** (1995), no. 1, 43–76.

[Rap78] M. Rapoport, *Compactifications de l'espace de modules de Hilbert-Blumenthal*, Compositio Math. **36** (1978), no. 3, 255–335 (French).

[Saw16] William F. Sawin, *Ordinary primes for Abelian surfaces*, C. R. Math. Acad. Sci. Paris **354** (2016), no. 6, 566–568, DOI 10.1016/j.crma.2016.01.025 (English, with English and French summaries). MR3494322

[Ser12] Jean-Pierre Serre, *Lectures on* $N_X(p)$, Chapman & Hall/CRC Research Notes in Mathematics, vol. 11, CRC Press, Boca Raton, FL, 2012.

[TZ16] Jesse Thorner and Asif Zaman, *A Chebotarev variant of the Brun-Titchmarsh theorem and bounds for the Lang-Trotter conjectures* (2016). Available on arXiv:1606.09238.

[vdG88] Gerard van der Geer, *Hilbert modular surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 16, Springer-Verlag, Berlin, 1988. MR930101

[Sch68] Wolfgang M. Schmidt, *Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height*, Duke Math. J. **35** (1968), 327–339.

[Yan10] Tonghai Yang, *An arithmetic intersection formula on Hilbert modular surfaces*, Amer. J. Math. **132** (2010), no. 5, 1275–1309.

[Zyw14] David Zywina, *The splitting of reductions of an abelian variety*, Int. Math. Res. Not. IMRN **18** (2014), 5042–5083. MR3264675