EQUIDISTRIBUTION OF CM POINTS ON SHIMURA CURVES AND TERNARY THETA SERIES

FRANCESCO MARIA SAETTONE

Abstract. We prove an equidistribution statement for the reduction of Galois orbits of CM points on the special fiber of a Shimura curve over a totally real field, considering both the split and the ramified case. This result is achieved by associating to the reduction of CM points certain Hilbert modular forms of weight $3/2$ and by analyzing their Fourier coefficients. Moreover, we also deduce the Shimura curves case of the integral version of the André–Oort conjecture.

Contents

1. Introduction

In [27] Jetchev and Kane proved an equidistribution result for a reduction of Galois orbits of Heegner points in a modular curve with both the conductor and the discriminant going to infinity. In this work we generalize their result to quaternionic Shimura curves over a totally real number field F, where we also allow the quaternion algebra both to be split and to ramify at the prime at which the reduction takes place.

Since the beginning of this century, the equidistribution of the reduction of Galois orbits of CM points has been a fruitful area of research. In [36], Michel gave a subconvexity bound for certain Rankin–Selberg L-functions and used this bound to prove an equidistribution property for Galois orbits of supersingular elliptic curves. His result was recently generalized to a simultaneous equidistribution in [1], exploiting dynamics and ergodic theory. On the other hand, Cornut and Vatsal in [9] proved that the reductions of CM points (modulo a non-split prime in a CM extension of F) of a Shimura curve are equidistributed in the supersingular locus. This was obtained by the mean of Ratner's theorem, with the noticeable application consisting of the Mazur's conjecture on the non-vanishing of central values of automorphic L-functions. In [41], we proved a companion equidistribution on the special fiber of a Shimura curve attached to some ramified primes. Moreover, in [49] a variant of these equidistributions for PEL Shimura varieties is treated. Lastly, Jetchev and Kane in [27] partially generalized [9] for modular curves.

For a overview on analogous recent p-adic equidistributions, we refer to [13]. In the rest of this introduction, we concisely state our main result, the idea of its proof, and a diophantine application.

Let B be an indefinite quaternion algebra over F. Let also K be a CM extension of F. In this work we deal we both the following situations:

(1) B is ramified at v and v ramifies in K;
(2) B is unramified at v and v is inert in K.

For the sake of clarity, we underline that in the setting of Shimura curves CM points are sometimes called Heegner points, in particular in the case of modular curves. We follow the terminology of CM point to differentiate between the points on the Shimura curve and the corresponding points on an elliptic curve, which we do not discuss.

Let $\{s_1, ..., s_h\}$ be the supersingular or superspecial locus[1] of the special fiber of a Shimura curve. The moduli interpretations of its local integral models at v allow us to consider the endomorphism ring $\mathrm{End}(s_i)$ of the supersingular point, and to define $w(s_i)$ as the number of units modulo $\{\pm 1\}$ of $\mathrm{End}(s_i)$. In the superspecial case, the analogous $w(s_i)$ is defined through the construction of Section 2.3.4.
Let $\star \in \{\mathrm{ss}, \mathrm{ssp}\}$, and define $\mu^\star$ be the normalized counting measure on the supersingular (respectively, superspecial) locus $\{s_1, ..., s_h\}$ of special fiber of the Shimura curve given by

$$\mu^\star(s_i) = \frac{w(s_i)^{-1}}{\sum_{j=1}^h w(s_j)^{-1}} \, .$$

Moreover, in the case of B_v ramified, let also $\mu_{\mathrm{in}}^{\mathrm{ssp}}$ be the normalized counting measure on the set of irreducible component $\{c_1, ..., c_k\}$, given by

$$\mu_{\mathrm{in}}^{\mathrm{ssp}}(c_i) = \frac{w(c_i)^{-1}}{\sum_{j=1}^h w(c_j)^{-1}}$$

where $w(c_i) = \# \mathrm{End}(c_i)/2$ is the *weight* of c_i. Here by $\mathrm{End}(c_i)$ we mean the endomorphisms of the formal group attached to the reduction of a CM point modulo an inert prime, which lands on the connected component c_i.

The main result of our work goes as follows.

Theorem A. The reductions at v of the Galois orbits of CM points are equidistributed for the discriminants and the conductors varying, i.e., for their absolute norms going to infinity:

- in the supersingular locus of the Shimura curve with respect of μ^{ss} for $B_v \simeq M_2(F_v)$ and v inert or ramified in K;
- in the superspecial locus of the Shimura curve with respect to μ^{ssp} for B_v ramified and v ramified in K;
- in the smooth locus of the Shimura curve with respect to $\mu_{\mathrm{in}}^{\mathrm{ssp}}$ for B_v ramified and v ramified in K.

This theorem gives a generalization of [9], [41] and, of course, of [27]. Stricto sensu, this generalization for [9] and [41] is only partial: in fact, although we do not fix the discriminant, we can only reduce by a single prime, loosing the simultaneous reduction of the two aforementioned works.

Let us briefly sketch the main objects and techniques we make use of.
We begin describing the special points of the indefinite and definite Shimura curves, both as adelic double quotients as in [47] and in terms of their moduli interpretations. In order to give these two descriptions also for their reduction in the special fiber at v, we turn to study the local integral models of the Shimura curve at v, for B_v unramified and ramified respectively. This changes dramatically their geometry and consequently the arithmetic of their moduli problems. In particular, in the ramified case we exploit the

[1] Note that these two sets do not have the same cardinality.

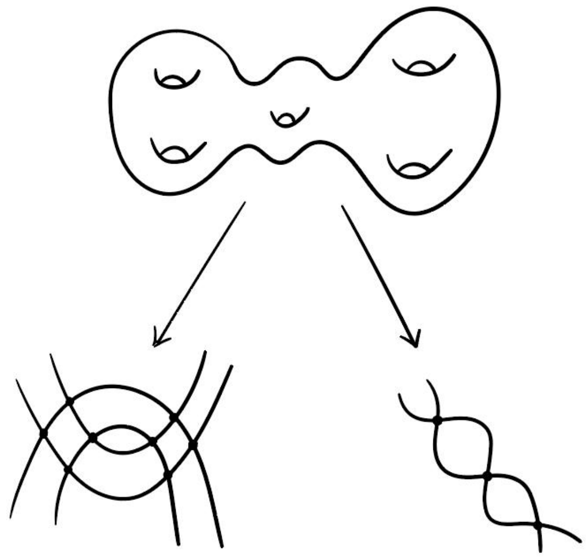

Figure 1. The illustration (drawn by Francesco Beccuti) represents the Shimura curve over $\mathbb{Q}$ of discriminant 77 and maximal level structure, whose CM points reduced modulo 7 land in the superspecial locus (on the left) and to the supersingular locus for a prime $p \neq 7, 11$. Note that on the right hand side the two curves should intersect 6 times, since the special fiber must have genus 5.

Cerednik–Drinfeld uniformisation as in [4] and the related moduli intepretation of the Drinfeld plane in terms of formal $\mathscr{O}_{B_v}$-modules. On the other hand, if $B_v \simeq M_2(F_v)$, only a local uniformisation in terms on Lubin–Tate spaces is available, and we essentially use the moduli interpretation of [7]. All of this eventually allows us to state a correspondence between the set of CM points x of fixed discriminant and conductor reducing to a supersingular or superspecial point s and the set of (conjugacy classes of) optimal embeddings $\mathrm{End}(x)$ into $\mathrm{End}(s)$. This correspondence has a "numerical" incarnation in the formula of Lemma 2.7.

Subsequently, we discuss Hilbert modular forms of half-integral weight and their automorphic counterparts, so to exploit an adelic setting. To a supersingular or superspecial point s, we associate an Eichler order R_s and then the so called Gross lattice, that is, a ternary quadratic form Q_s attached to s. Since at least Jacobi, quadratic forms and automorphic forms are allied subjects, and in fact Q_s gives rise to a theta series of weight 3/2. Such theta series were introduced in Gross' seminal paper [23], and they formed already a key idea in [15] and similarly in [27] where the authors exploited a subconvexity bound on their Fourier coefficients. Now, the set of the optimal embeddings previously introduced is in correspondence with primitive representations of of the discriminant of x by (the adelization of) Q_s. Next step, which was the main technical novelty of [27], consists in the analysis of the Whittaker–Fourier coefficients of the (automorphic) theta series attached to the genus and the spinor genus of Q_s. In particular, we prove that, under certain conditions, the genus and spinor genus mass, as defined in (3.8), are equal. To prove this, we make use of some class number formula and of Vatsal's equidistribution in Lemma 3.16. Moreover, we also need to adapt to totally real fields some properties of those theta series. To do so, we decided to work in an autormorphic setting. We thus exploit the beautiful papers of Gelbart–Piatetsky-Shapiro [21] and [22]: these works develop an automorphic, representation theoretic Shimura correspondence, which can be viewed as a first case of the theta-correspondence.

The last crucial ingredient for Theorem A follows from the subconvexity bound of [6] on the Whittaker–Fourier coefficients of cuspidal automorphic forms of weight 3/2. The ineffectivity of our result is due to Brauer–Siegel's lower bound for class numbers. However, this issue might be solved by assuming the Generalized Riemann Hypothesis or by the (weaker assumption of the) non-existence of Landau–Siegel zeros. Let $F = \mathbb{Q}$ and D be the discriminant of an imaginary quadratic extension of $\mathbb{Q}$ with class number $h(D)$. Then by one of the two assumptions above Hecke showed there is an effective constant $c > 0$ such that $h(D) > c\sqrt{|D|}(\log |D|)^{-1}$.

Lastly, in Section 4, as a complement of our equidistribution result, we deduce the integral version of the André–Oort conjecture in arithmetic pencil introduced in [39] in the case of Shimura curves over $\mathbb{Q}$. While the classical André–Oort conjecture concerns the Zariski topology of Shimura varieties, this variant considers their integral models, which are defined over $\operatorname{Spec}(\mathscr{O}_E)$, where E is their reflex field, hence the pencil over the arithmetic base $\operatorname{Spec}(\mathscr{O}_E)$.

The strategy closely follows the lines of [40], where the case of the modular curve $Y(1)$ over $\mathbb{Z}$ is established.

1.1. **Notation and Conventions.** We list some notations used throughout this work:

- we denote by $\mathbb{H}$ the classical Hamilton's quaternions;
- $\mathbb{A}$ denotes the ring of adeles over $\mathbb{Q}$, and $\mathbb{A}_f$ the finite adeles;
- for a totally real number field F, let $\mathscr{O}_F^+ := \mathscr{O}_F \cap F^+$;
- for a quaternion algebra B over F, we denote by $\operatorname{Ram}(B)$, the set of places of F where B ramifies; the same symbol, with f or ∞ subscript, denotes the restriction to finite or archimedean places respectively;
- for a local field F_v, we denote by $\breve{F}_v$ the completion of the maximal unramified extension;
- let G_K^{ab} denote the abelianization of the absolute Galois group of K;
- for a set X, we denote by $\mathcal{C}(X, \mathbb{C})$ the set of continuous functions on it;
- for a group scheme G, we denote its automorphic quotient by $[G] := G(\mathbb{A})/G(F)$;
- we denote by N the absolute norm of a number field, while nr denotes the reduced norm of a quaternion algebra;
- for any two functions f, g, we use the notations $f \ll g$ and $f = O(g)$ interchangeably.

Acknowledgments. It is a pleasure to thank Professor Daniel Disegni for his guidance with my doctoral thesis, which this article is part of. I am also grateful to Zev Rosengarten for useful conversations and correspondence. This work was supported by ISF grant 1963/20 and BSF grant 201825.

2. Shimura Curves

2.1. **Quaternion Algebras and Ramification.** For a totally real number field F of degree d over $\mathbb{Q}$, let B be a quaternion algebra over F. In this section, after imposing some conditions on the ramification of the quaternion algebra B at some places of F, we associate two curves to B, according to the fact that B satisfies one of the following conditions:

- there is a unique real place v of F such that $B_v = B \otimes F_v \simeq M_2(F_v)$, i.e., B is *indefinite*;
- for every real place B_v is non-split, i.e., B is *definite*.

2.1.1. *Ramification Setting.* From now on, let K denote a CM quadratic extension of F, and suppose that B is split by K, i.e., K_v is a field for every finite place of F where B ramifies. Fix an embedding $\rho\colon K \hookrightarrow B$ over F. Let also v be a finite place of F. In what follows we consider the following two situations:

(1) B is ramified at v and v ramifies in K;
(2) B is unramified at v and v is inert in K.

Indeed, this does not rule all the possible cases out. For instance, in (1), the prime v could also be inert. The geometric interpretation of this case is described in [41, Prop.2.13]. At any rate, at the level of the equidistribution both possibilities in (1) would proceed in parallel, so we decided to focus exclusively on v ramifying in K.

2.1.2. *Indefinite Case.* We first consider B indefinite at one archimedean place of F, which we denote by $\tau_1\colon F \hookrightarrow \mathbb{R}$. Then we can fix an isomorphism $B \otimes \mathbb{R} \simeq M_2(\mathbb{R}) \oplus \mathbb{H}^{d-1}$. This isomorphism induces a map $B^\times \to \mathrm{GL}_2(\mathbb{R})$ which gives an action of $B^\times$ on the conjugacy class of h_0, which is isomorphic to $\mathscr{H} = \mathbb{C} - \mathbb{R}$, by Möbius transformations.

Let G be the reductive group over $\mathbb{Q}$ whose functor of points sends a commutative $\mathbb{Q}$-algebra A to

$$G(A) = (B \otimes_{\mathbb{Q}} A)^\times. \tag{2.1}$$

This means that $G = \mathrm{Res}_{F/\mathbb{Q}} B^\times$ and there is a real embedding $h_0\colon \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\mathbb{G}_m \to G_{\mathbb{R}}$ with trivial coordinates at τ_i for $i \geqslant 2$.

For any open subgroup U of $G(\mathbb{A}_f)$ which is compact modulo $\widehat{F}^\times$, we consider

$$\mathcal{S}_U^{\mathrm{an}} := G(\mathbb{Q})\backslash \mathscr{H} \times G(\mathbb{A}_f)/U \cup \{\mathrm{cusps}\} \tag{2.2}$$

whose canonical model is the Shimura curve $\mathcal{S}_U$ over F. It is a proper and smooth curve over its reflex field F. Note that the set $\{\mathrm{cusps}\}$ is non-empty if and only if $B = M_2(\mathbb{Q})$. If $F = \mathbb{Q}$, then $\mathcal{S}_U$ is a coarse moduli space of QM abelian surfaces.

On the other hand, if $d > 1$, Shimura curves do not have a natural moduli interpretation. Despite so, by the work of Carayol [7], we have that $\mathcal{S}_U$ has a finite map to another Shimura curve $\mathcal{S}_{U'}$ which, if U' is small enough, is a moduli space of QM abelian varieties with a U'-level structure and a polarization both compatible with the quaternionic multiplication (see [48, Prop.1.1.5]). Since $\mathscr{S}'_{U'_v}$ is a fine moduli scheme for such abelian schemes over schemes with level structures, there is a universal object of its moduli problem, called *universal* abelian surface and denoted by $\mathscr{A} \to \mathscr{S}'_{U'_v}$. For each geometric point $x = \mathrm{Spec}(E)$ of $\mathscr{A}$, the fiber $\mathscr{A}_x$ is a QM abelian surface as above structure defined over E.

The moduli interpretation yields the remarkable consequence that the Shimura curve has a proper[2] regular integral model $\mathscr{S}_U$ over $\mathscr{O}_F$.

2.1.3. *Definite Case.* Let us now deal with a definite quaternion algebra. For the rest of this work fix a finite prime ℓ of F such that $B_\ell \simeq M_2(F_\ell)$. Consider an Eichler order R of level $\mathfrak{n}$. An *orientation* on R consists of a morphism $\mathfrak{o}_\star\colon R \otimes k_\ell \to k_\star$, for $\star \in \{\ell, \ell^2\}$, where $k_\star = k_\ell$ if $\ell | \mathfrak{n}$ and $k_\star = k_{\ell^2}$ otherwise. We refer to [5, Sec.1.1] for an exhaustive description of the orientations.

We denote by $\mathrm{Cl}(B)$ the set of all conjugacy classes of oriented Eichler orders of level $\mathfrak{n}$ in B, and we recall that it can be viewed adelically via the following bijection

$$\mathrm{Cl}(B) \simeq G(\mathbb{Q})\backslash G(\mathbb{A}_f)/\widehat{R}^\times, \tag{2.3}$$

where $\widehat{R}^\times$ denotes the adelization of R and G is the reductive group defined in (2.1) adapted to the definite setting. By strong approximation in $G(\mathbb{A}_f)$, it follows that $\mathrm{Cl}(B)$ is finite.

Following [23, p.131], we now attach to B an algebraic curve. Let $\mathscr{P}$ be the conic curve over $\mathbb{Q}$ whose functor of points send a commutative $\mathbb{Q}$-algebra A to

$$\mathscr{P}(A) = \{x \in B \otimes A : x \neq 0,\ nr(x) = tr(x) = 0\}/A^\times.$$

We define the *Gross curve*[3] of level R as

$$X_R = G(\mathbb{Q})\backslash \mathscr{P} \times G(\mathbb{A}_f)/\widehat{R}^\times,$$

where $G(\mathbb{Q}) = \mathrm{Aut}(\mathscr{P})$ acts by conjugation on $\mathscr{P}$. Given representatives $(g_i)_{i=1}^r$ of $\mathrm{Cl}(B)$ via (2.3), we define $\Gamma_i := g_i \widehat{R} g_i^{-1} \cap G(\mathbb{Q})$ is a finite subgroup of $G(\mathbb{Q})$. We thus obtain a collection of genus zero conic

[2]In the modular curves case one needs to add finitely many cusps for properness.

[3]Also known as *definite* Shimura curve, as in [5, p.420].

curves[4] $Y_i = \Gamma_i \backslash \mathscr{P}$ defined over $\mathbb{Q}$, and we write the Gross curve as

$$X_R = \bigsqcup_{i=1}^{r} Y_i.$$

The set of K-rational points of the conic curve $\mathscr{P}$ is identified with $\mathrm{Hom}(K,B)$, as explained in [23, p.131]. By this identification, we obtain that

$$X_R(K) = \{(f,[R]) : f \in \mathrm{Hom}(K,B),\ [R] \in \mathrm{Cl}(B)\}.$$

Lastly, since the connected components Y_i have genus zero, the Picard $\mathrm{Pic}(X_R)$ is a free $\mathscr{O}_F$-module with a basis indexed by $\mathrm{Cl}(B)$.

2.2. **Special Points and Quaternion Algebras.**

2.2.1. *Indefinite Case.* For any $F \otimes \mathbb{A}_f$-embedding $\tau \colon K \otimes \mathbb{A}_f \hookrightarrow B \otimes \mathbb{A}_f$, the group $\tau(K^\times)$ acts on the Shimura curve $\mathcal{S}_U$. We thus define the scheme of *CM points* by (K,τ) as the fixed-point (affine) subscheme $\mathcal{S}_U^{\tau(K^\times)}$.
In other words, a point z of $\mathcal{S}_U$ is a CM point by K if it can be represented by $(z_0, g) \in \mathscr{H} \times G(\mathbb{A}_f)$ via (2.2), where z_0 is the unique point fixed by $K^\times$. By the work of Shimura, it is a finite subscheme of $\mathcal{S}_U$ defined over K^{ab}. By taking the union of $\mathcal{S}_U^{\tau(K^\times)}$ over all pairs (K,τ), we obtain the CM ind-subscheme of $\mathcal{S}_U^{\mathrm{CM}}$. The absolute Galois group of K, which we denote by G_K, acts on $\mathcal{S}_U^{\mathrm{CM}}$ via

$$\sigma.(x_0, g) = (x_0, \mathrm{rec}_K(\sigma) g)$$

where rec_K is Artin's reciprocity map. If we consider CM points of conductor c, this action factors through $\mathrm{Gal}(H[c], K)$, where $H[c]$ is the ring class field of K of conductor c.

Consider now an order R of B of type[5] $(\mathfrak{n}, K)$ as constructed in [48, Sec.1.5.1] and the corresponding Shimura curve of level $\widehat{F}^\times \widehat{R}^\times$. Then for z a CM point by K we consider

$$\mathrm{End}(z) := g\widehat{R}^\times g^{-1} \cap \rho(K)$$

which is an order in $K = \rho(K)$ independent of the choice of $g \in G(\mathbb{A}_f)$. The *conductor* of z is defined as the unique $\mathscr{O}_F$-ideal such that

$$\mathrm{End}(z) = \mathscr{O}_F + c\mathscr{O}_K.$$

Moreover, the discriminant of $\mathrm{End}(z)$ is of the form Dc^2, where D is the discriminant of K relative to F and c the conductor.

We say that a CM point z corresponding to (z_0, g) has an *orientation* if, for a finite prime v coprime with c, the morphism $g^{-1}\rho g$ is $R_v^\times$-conjugated to ρ in $\mathrm{Hom}(\mathscr{O}_{K_v}, R_v)/R_v^\times$.

Lemma 2.1. *Let c be coprime with $\mathfrak{n}$. Then there are $2^{\omega(\mathfrak{n})}$ Galois orbits of CM points of conductor c, where $\omega(\mathfrak{n})$ is the number of prime factors of $\mathfrak{n}$. Moreover, every such orbit has cardinality equal to $\#\,\mathrm{Pic}(\mathscr{O}_c)$.*

Proof. Let us recall that the Shimura curve of level $\widehat{R}^\times$ has an action by the Atkin–Lehner group

$$\mathcal{W} = \{b \in G(\mathbb{A}_f) : b^{-1}\widehat{R}^\times b = \widehat{R}^\times\}/\widehat{R}^\times$$

which has 2^s elements, for s the number of prime factors of $\mathfrak{n}$. This yields an action of $\mathcal{W}$ on $\mathcal{S}_U^{\mathrm{CM}}$ which preserves the conductor. By the work of Gross [23], the action of $\mathrm{Pic}(\mathscr{O}_c) \times \mathcal{W}$ on the set of CM points of conductor c is free and transitive and the $\mathrm{Pic}(\mathscr{O}_c)$-orbits of CM points correspond to sets of oriented CM points. This implies that there are $2^{\omega(\mathfrak{n})} \cdot \#\,\mathrm{Pic}(\mathscr{O}_c)$ such points of conductor c, and so we conclude. □

[4]Which are in general not isomorphic to $\mathbb{P}^1$.
[5]I.e., an order of discriminant $\mathfrak{n}$ which contains $\rho(\mathscr{O}_K)$.

Next description characterizes CM points in terms of an adelic double quotient. Let T be the $\mathbb{Q}$-rational torus in G.

Lemma 2.2. *The set of CM points in $\mathcal{S}_U$ is in bijection with*

$$T(\mathbb{Q})\backslash G(\mathbb{A}_f)/U.$$

Proof. See [47, 5.2.5.2.2]. □

2.2.2. *Definite Case.* Here we consider the definite case, i.e., special points living on some Gross curve, whose theory is considerably less convoluted than in the indefinite case.

For any order $\mathscr{O}$ of K, we fix an orientation on it by choosing a morphism $\mathscr{O}_\ell \to k_\star$ where $k_\star$ is as in Section 2.1.3.

A *Gross* point of conductor c consists of a pair (f, R) where $f\colon K \to R$ is an oriented optimal embedding, which means that $f(K) \cap R = f(\mathscr{O}_c)$ where $\mathscr{O}_c$ is taken up to conjugation by $G(\mathbb{Q})$. Equivalently, Gross points are the image of

$$\mathscr{P}(K) \times G(\mathbb{A}_f)/\widehat{R}^\times$$

in $X_R(K)$. This immediately implies the K-rationality of Gross points. Let us denote by $\mathrm{Gr}(c)$ the set of Gross points of conductor c. Moreover, a Gross point (f, R) represented by $(x, g) \in \mathscr{P} \times G(\mathbb{A}_f)$ has discriminant D if and only if

$$f(K) \cap g\widehat{R}g^{-1} = f(\mathscr{O})$$

is the image of the order of discriminant D. Thus, Gross points of discriminant D correspond to equivalence classes modulo $R_i^\times$ of optimal embedding of $\mathscr{O} \hookrightarrow R_i$ in each component Y_i. We denote the cardinality of the set of such equivalence classes by $h(\mathscr{O}, R_i)$.

Set T to be the $\mathbb{Q}$-rational torus in G. For any Gross point $(f, R) \in \mathrm{Gr}(c)$ the optimal embedding f induces, by scalar extension, a map $T(\mathbb{A}) \to G(\mathbb{A})$. This gives an action of the Picard group of $\mathscr{O}_c$

$$\mathrm{Pic}(\mathscr{O}_c) = T(\mathbb{Q})\backslash T(\mathbb{A})/\widehat{\mathscr{O}}_c^\times$$

on $\mathrm{Gr}(c)$. By [23, p.133], this action is simply transitive.

Next result recalls how to describe Gross points as an adelic double quotient.

Lemma 2.3. *The set of Gross points in X_R is in bijection with*

$$T(\mathbb{Q})\backslash G'(\mathbb{A}_f)/\widehat{R}^\times.$$

Proof. See [27, Lemma 2.2]. □

2.3. **Reductions of Integral Models.** In this section we consider the indefinite quaternion algebra B exclusively.

2.3.1. *Universal v-Divisible Group.* Let A be a complete Noetherian local ring of residue characteristic p, and let v be a p-adic place of F. A v-divisible[6] group $\mathscr{G}$ of height h over A is the colimit of a tower of affine, finite flat group schemes $(\mathscr{G}_n)_{n\geqslant 1}$ over A of order p^{nh} such that $\mathscr{G}_n = \mathscr{G}_{n+1}[v^n]$. Taking the connected and étale parts[7] of $\mathscr{G}$, denoted by $\mathscr{G}^\circ$ and $\mathscr{G}^{\text{ét}}$, there is a short exact sequence

$$0 \to \mathscr{G}^\circ \to \mathscr{G} \to \mathscr{G}^{\text{ét}} \to 0$$

which splits if A is a perfect field. Moreover, let us recall that a *formal* group is a group object in the category of formal schemes. Then there is an equivalence of categories between connected v-divisible groups and formal groups.

We now assume that the level structure decomposes in the form $U = U^v \cdot U_v$ where U^v is sufficiently small and $U_v \simeq \mathscr{O}_{B_v}^\times$, namely the level structure is maximal at v.

[6]Also known as *Barsotti–Tate*.

[7]Defined just by the of thelatter group schemes $\mathscr{G}_n$.

The curve $\mathscr{S}_{U_v}$ carries a universal v-divisible $\mathscr{O}_{B_v}$-module $\mathscr{G}$ which comes from the v-divisible group $\mathscr{A}[v^\infty]$ by taking the pull-back of $\mathscr{A}$ via the finite map $\mathscr{S}_{U_v} \to \mathscr{S}'_{U'_v}$. To describe $\mathscr{G}$, choose an auxiliary quadratic field F' as in [48, p.33] which is split at v and fix an isomorphism $f\colon \mathscr{O}_{F'_v} \simeq \mathscr{O}_{F_v} \oplus \mathscr{O}_{F_v}$. We therefore define

$$\mathscr{G} := f^{-1}(0,1)\mathscr{A}[v^\infty].$$

Let $\mathscr{O}_{\mathrm{un}}$ be any unramified quadratic extension of $\mathscr{O}_{F_v}$ contained in $\mathscr{O}_{B_v}$. By [48, Prop.1.2.4], we have that $\mathscr{G}$ is a *special* v-divisible $\mathscr{O}_{B_v}$-module, i.e., the induced action of $\mathscr{O}_{B_v}$ on $\mathrm{Lie}(\mathscr{G}) := \mathrm{Lie}(\mathscr{G}^\circ)$ makes $\mathrm{Lie}(\mathscr{G})$ a locally free sheaf over $\mathscr{O}_{\mathscr{S}} \otimes_{\mathscr{O}_{F_v}} \mathscr{O}_{\mathrm{un}}$ of rank 1. Moreover, $\mathscr{G}$ is étale away from v.

2.3.2. *Unramified Case.* Assume that B is unramified at v, and fix an isomorphism $j\colon \mathscr{O}_{B_v} \simeq M_2(\mathscr{O}_v)$. Let also be the level structure U as in Section 2.3.1. By [7] (and by [32, Chapt.13] if $F = \mathbb{Q}$), we have that $\mathscr{S}_{U_v}$ has good reduction.

Let s be a geometric point of the special fiber of $\mathscr{S}_U$ at v. Since we have fixed j, every $\mathscr{O}_{B_v}$-module M can be uniquely decomposed via idempotents as

$$M = M^1 \oplus M^2 := \begin{bmatrix} 1 & \\ & \end{bmatrix} M \oplus \begin{bmatrix} & \\ & 1 \end{bmatrix} M$$

where the $\mathscr{O}_{F_v}$-modules M^1 and M^2 are isomorphic. Therefore we can write $\mathscr{G}$ as

$$\mathscr{G} = \mathscr{G}^1 \oplus \mathscr{G}^2$$

with the summands are isomorphic as v-divisible $\mathscr{O}_{F_v}$-modules.

Let $\underline{F_v/\mathscr{O}_{F_v}}$ denote the constant v-divisible group of height 1 given by the colimit of $(\underline{v^{-n}\mathscr{O}_{F_v}/\mathscr{O}_{F_v}})$, which is étale. Since the v-divisible $\mathscr{O}_{F_v}$-modules $\mathscr{G}^i_s$ have height 1 and dimension 2, they are isomorphic to one of the following objects:

(1) the direct sum $X_1 \oplus \underline{F_v/\mathscr{O}_{F_v}}$, where X_1 is the unique[8] formal $\mathscr{O}_{F_v}$-module of height 1, so that s is called *ordinary*;
(2) to the unique formal $\mathscr{O}_{F_v}$-module of height 2 and dimension 1, so that s is called *supersingular*.

In other words, $\mathscr{G}^i_s$ are supersingular if $\mathscr{G}^i_s = \mathscr{G}^\circ$, while they are ordinary if the étale part $\mathscr{G}^{\text{ét}}$ is non-trivial.

Let k be the algebraic closure of the residue field of F at v. We denote by $\mathscr{S}^{\mathrm{ss}}_{U,k}$ the finite étale subscheme of supersingular points over $\mathrm{Spec}(k)$. We will also refer to it as the *supersingular* locus. The complement of this finite set of points is naturally called the *ordinary* locus.

Let us recall that supersingular points are mutually isogenous, ad that, given such a point s, its endomorphism ring $\mathrm{End}(s) \otimes F$ is the quaternion algebra B' obtained by switching invariants at τ and v in B.

Geometrically, the irreducible components of special fiber $\mathscr{S}_{U,k}$ are smooth connected curves[9] and they intersect each other transversally only in the supersingular locus, so that supersingular points corresponds to the singularities of the special fiber. We refer to [7, 9.4.4] for more details.

Next Lemma gives an adelic description of the supersingular points.

Lemma 2.4. *The set of supersingular points in $\mathscr{S}_{U,k}$ is in bijection with*

$$G'(\mathbb{Q})\backslash G'(\mathbb{A}_f)/U',$$

where $U' = \mathscr{O}^\times_{B'_v} \cdot U^v$.

Proof. See [47, p.259]. ◻

[8] Therefore X_1 coincides with the geometric generic fiber of the Lubin-Tate formal group $\mathscr{L}\mathscr{T}$, i.e., $\mathscr{L}\mathscr{T}|_{\mathrm{Spec}(\overline{F}_v)}$.

[9] Never isomorphic to $\mathbb{P}^1$.

2.3.3. *Ramified case.* Assume that B is ramified at v and that U_v is maximal. In this setting, by the work of Drinfeld [14] we have that $\mathscr{G} = \mathscr{G}^\circ$, i.e., $\mathscr{G}$ is a formal group.

Let B' be the quaternion algebra obtained by changing the invariant of B at v and τ, i.e., B' is definite and unramified at v. Consequently, we denote by G' the reductive group whose functor of points is given by $G'(A) = (B \otimes_{\mathbb{Q}} A)^\times$, for a $\mathbb{Q}$-algebra A. We then fix an isomorphism $G'(\mathbb{A}_f) \simeq M_2(F_v) \cdot G(\mathbb{A}_f^v)$.

Consider the formal scheme $\widehat{\Omega}$ over $\mathscr{O}_{F_v}$ obtained by successive blow-ups of rational points on the special fiber over k of $\mathbb{P}^1$, whose generic fiber is the rigid-analytic Drinfeld plane Ω over F_v such that

$$\Omega(\mathbb{C}_v) = \mathbb{P}^1(\mathbb{C}_v) - \mathbb{P}^1(F_v).$$

We denote by $\widehat{\mathscr{S}_U}$ the completion of $\mathscr{S}_U$ along its special fiber at v, and assume that U_v is maximal, i.e., $U_v \simeq \mathscr{O}_{B_v}^\times$ and that U^v is sufficiently small. A wonderful result of Cerednik and Drinfeld gives the following uniformisation

$$\widehat{\mathscr{S}_U} \simeq G'(\mathbb{Q}) \backslash (\widehat{\Omega} \widehat{\otimes} \breve{\mathscr{O}}_{F_v}) \times \mathbb{Z} \times G'(\mathbb{A}_f^v)/U. \tag{2.4}$$

For a detailed discussion of the Cerednik-Drinfeld uniformisation, we invite the reader to see [3].

Let $\mathbb{X}$ be the unique (up to isogeny) formal module of height 4 over k_v. By Drinfeld moduli interpretation [3, p.107], $\widehat{\Omega} \widehat{\otimes} \breve{\mathscr{O}}_{F_v}$ can be viewed as the moduli space of the (isomorphism classed of the) following objects:

- a formal special $\mathscr{O}_{B_v}$-module X of height 4;
- a height zero quasi-isogeny $\varrho \colon \mathbb{X} \to X_{k_v}$.

Note that, since this moduli interpretation is proved as usual via the representability of a moduli problem, this quasi-isogeny excludes a stacky situation. For a detailed description of the special fiber, see [41].

Let s be a geometric point of the special fiber of $\mathscr{S}_U$ at v. Then by (2.4) and Drinfeld moduli interpretation, s corresponds to a formal $\mathscr{O}_{B_v}$-module X_v. Then s is *supersingular* if X_v is isogenous to the direct sum of two formal $\mathscr{O}_{F_v}$-modules Y_v of height 2 and dimension 1 such that $\mathrm{End}(Y_v) \simeq \mathscr{O}_{B_v}$. It is not difficult to see, as in [41, Lemma 2.9], that every geometric point in this special fiber is supersingular. Nonetheless, there is a distinguished set of points encoding arithmetic and geometric properties in a manner similar to the supersingular locus in the unramified case. A geometric point s is called *superspecial* if the associated formal $\mathscr{O}_{B_v}$-module X_v is isomorphic to $Y_v \oplus Y_v$; this isomorphism is unique up to $\mathrm{GL}_2(\mathscr{O}_{B_v})$-conjugation. For such a point s, its endomorphism ring is $\mathrm{End}(X_v) \simeq M_2(\mathscr{O}_{B_v})$. The $\mathscr{O}_{B_v}$-action on X is given by

$$\iota \colon \mathscr{O}_{B_v} \hookrightarrow \mathrm{End}(X_v) \simeq M_2(\mathscr{O}_{B_v}).$$

The finite subscheme $\widehat{\mathscr{S}}_{U,k}^{\mathrm{ssp}}$ over $\mathrm{Spec}(k)$ consisting of superspecial points is thus called the *superspecial* locus.

Geometrically, the irreducible components of the special fiber of $\widehat{\Omega}$ are projective lines $\mathbb{P}_k^1$ intersecting each other in the superspecial locus, which does is in bijection with the ordinary double points of $\widehat{\mathscr{S}_{U,k}}$. Thus the complement of these singularities is the *smooth* locus of $\widehat{\mathscr{S}_{U,k}}$, which we denote by $\mathscr{S}_{U,k}^{\mathrm{sm}}$. For more details on this smooth locus, see [41, Sec.2].

As for the supersingular points in Lemma 2.4, also superspecial points admit an adelic interpretation.

Lemma 2.5. *We have that the set of superspecial points on $\widehat{\mathscr{S}_{U,k}}$ corresponding to the class of a fixed ι are in bijection with*

$$G'(\mathbb{Q})_0 \backslash G'(\mathbb{A}_f^v)/U^v$$

where $G'(\mathbb{Q})_0$ are the elements in the centralizer of $\iota(\mathscr{O}_{B_v})$.

Proof. See [47, Lemma 5.4.5]. □

2.3.4. *Eichler Orders associated to Supersingular and Superspecial Points.* Next construction is a variant of [23, pp.171-172] including the ramified setting.

Let v be a non-archimedean place of F and consider the special fiber $\mathscr{S}_v$. Let also $x \in \mathcal{S}^{\mathrm{CM}}$. As usual, we consider s, namely the reduction of x modulo v, in the following cases:

(1) v unramified in B and inert in K;
(2) v ramified in B and ramified in K.

In the unramified case, i.e., for s in the supersingular locus, then we set $R_s = \mathrm{End}(s)$ as in [27, Sec.4.1]. On the other hand, if v is ramified in B, i.e., for s in the superspecial locus, we construct the desired Eichler order R_s as follows.

Let us consider now the ramified case with s superspecial. Let X_k correspond to the superspecial point s in $\widehat{\Omega}_k$, whose quaternionic action is given by $\iota\colon \mathscr{O}_{B_v} \hookrightarrow M_2(\mathscr{O}_{B_v})$. Consider the maximal orders

$$R_{\iota,v} := \mathrm{End}_{\iota(\mathscr{O}_{B_v})}(\mathbb{X}) \subset M_2(\mathscr{O}_{F_v}) = \mathrm{End}_{\mathscr{O}_{B_v}}(X)$$

where $M_2(\mathscr{O}_{F_v})$ has rank 4, and

$$\widehat{R}_f^v \subset \widehat{B}_f^v := B \otimes \mathbb{A}_{F,f}^v.$$

Let $\mathbb{B}'$ be the coherent[10] quaternion algebra over $\mathbb{A}_{F,f}$ and consider the Eichler order

$$\widehat{R} := R_{\iota,v} \times \widehat{R}_f^v \subset \mathbb{B}'.$$

There exists a unique quaternion algebra over B' such that $\mathbb{B}' \simeq B' \otimes \mathbb{A}_{F,f}$ and $\widehat{R} \subset B' \otimes \mathbb{A}_{F,f}$ so that

$$R_s := R \subset B'$$

and B' is ramified at τ and *not* ramified at v. Note that this change of invariants takes place at $R_{\iota,v}$, the centralizer of the action of $\iota(\mathscr{O}_{B_v})$.

2.3.5. *The Reduction Map.* Let $\bar{v}$ be a place of K^{ab} above v, and denote by $\mathscr{O}_{\bar{v}}$ and $k_{\bar{v}}$ its ring of integers at $\bar{v}$ and its residue field respectively.

The Shimura curve $\mathscr{S}_U$ is proper over $\mathrm{Spec}(\mathscr{O}_{F_v})$ both in the unramified and ramified cases, and so by the valuative criterion for properness we have that

$$\mathcal{S}_U(K^{\mathrm{ab}}) = \mathscr{S}_U(K^{\mathrm{ab}}) \simeq \mathscr{S}_U(\mathscr{O}_{\bar{v}}) \to \mathscr{S}_U(k_{\bar{v}}).$$

Since CM points are defined over K^{ab}, we obtain the following reduction map

$$red_v\colon \mathscr{S}_U^{\mathrm{CM}} \longrightarrow \mathscr{S}_U(k_{\bar{v}}).$$

Lemma 2.6. *The reduction of CM points in the special fiber modulo v lies in*

(1) *the supersingular locus, if B_v is unramified and v is inert or ramified in K;*
(2) *the superspecial locus, if B_v is ramified and v ramifies in K;*
(3) *in the smooth locus, if B_v is ramified and v is inert in K.*

In symbols,

$$red_v(\mathcal{S}^{CM}) \subseteq \begin{cases} \mathscr{S}_k^{ss}, & \text{for } B_v \simeq M_2(F_v), v \text{ inert or ramified in } K; \\ \mathscr{S}_k^{ssp}, & \text{for } B_v \text{ ramified}, \;\; v \text{ ramified in } \; K; \\ \mathscr{S}_k^{sm}, & \text{for } B_v \text{ ramified}, \;\; v \text{ inert in } \; K. \end{cases}$$

Proof. See [9, Lemma 3.1] for $B_v \simeq M_v(F_v)$ and [41, Prop.2.17] for the ramified case. □

We now conclude by writing down the three sequences of measures whose limit will give us our equidistribution result.

Let the geometric point s in the special fiber at v lie in

- the supersingular locus, for $B_v \simeq M_2(F_v)$;
- in the superspecial locus, for B_v ramified.

[10] This terminology was introduced in [46, p.3].

For $\star \in \{\mathrm{ss}, \mathrm{ssp}\}$, we define the following probability measures

$$\mu^{\star}_{D,c}(s) = \frac{1}{\#\Gamma_{D,c}} \sum_{\substack{x \in \Gamma_{D,c} \\ red_v(x) = s}} \mathbf{1}_s(x), \tag{2.5}$$

where $\mu^{\mathrm{ss}}_{D,c}$ is defined over $\mathscr{S}^{\mathrm{ss}}_{U,k}$ and $\mu^{\mathrm{ssp}}_{D,c}$ over $\widehat{\mathscr{S}}^{\mathrm{ssp}}_{U,k}$.

Lastly, let s be a geometric point in the special fiber at v, for B_v ramified, whose corresponding formal $\mathscr{O}_{B_v}$-module is *not* superspecial, so that it is supersingular, and it lies on one of the components $\{c_1, \dots, c_n\}$ of the smooth locus $\mathscr{S}^{\mathrm{sm}}_{U,k}$. Exactly as above in (2.5), we then obtain the analogous measure $\mu^{\mathrm{ssp}}_{\mathrm{in},D,c}$.

2.3.6. *Liftings of Reductions Maps.* Consider the two following maps

$$\pi^{\star} \colon \mathrm{Gr} \to \mathscr{S}^{\star}_k$$

where once again $\star \in \{\mathrm{ss}, \mathrm{ssp}\}$. By Lemmata 2.3, 2.4 and 2.5, the $\pi^{\star}$'s are the natural projections defined by

$$\begin{array}{ccc} & T(\mathbb{Q})\backslash G'(\mathbb{A}_f)/\widehat{R'}^{\times} & \\ {}^{\pi^{\mathrm{ss}}}\swarrow & & \searrow^{\pi^{\mathrm{ssp}}} \\ G'(\mathbb{Q})\backslash G'(\mathbb{A}_f)/\widehat{R'}^{\times} & & G(\mathbb{Q})_0\backslash G'(\mathbb{A}_f)/\widehat{R'}^{v,\times}. \end{array}$$

The following construction allows to move from the reduction maps to the projections $\pi^{\star}$.
For B_v unramified, by [8, Sec.3.2] there is a G_K-equivariant lifting θ_v such that the following diagram

$$\begin{array}{ccc} \mathcal{S}^{\mathrm{CM}} & \overset{\theta_v}{\dashrightarrow} & \mathrm{Gr} \\ {}_{red_v}\searrow & & \swarrow_{\pi^{\star}} \\ & \mathscr{S}^{\star}_k & \end{array}$$

is commutative. On the other hand, for B_v ramified, then the construction of the lifting θ_v follows almost verbatim the unramified case.

Let now R be an Eichler $\mathscr{O}_F$-order in B given by the maximal orders R' and R'', and consider the Shimura curve of level $\widehat{F}^{\times}\widehat{R}^{\times}$. By the interpretation of CM points as an adelic doubly quotients in Lemma 2.2 we obtain the two maps

$$\begin{array}{ccc} & T(\mathbb{Q})\backslash G(\mathbb{A}_f)/\widehat{R}^{\times} & \\ \swarrow & & \searrow \\ T(\mathbb{Q})\backslash G(\mathbb{A}_f)/\widehat{R'}^{\times} & & T(\mathbb{Q})\backslash G(\mathbb{A}_f)/\widehat{R''}^{\times}. \end{array}$$

Given a $z \in \mathcal{S}^{\mathrm{CM}}$, we denote its image under these maps as z' and z'' respectively. For $z = [g]$, then $K \cap g\widehat{R'}^{\times}g^{-1}$ is an $\mathscr{O}_F$-order in K whose conductor, denoted $c(z')$, is an integral $\mathscr{O}_F$-ideal coprime to $\mathrm{Ram}_f(B)$. Symmetrically, $c(z'')$ is the conductor coming from R''. We thus obtain the (*coarse*) *conductor* map

$$\mathbf{c} \colon z \in \mathcal{S}^{\mathrm{CM}} \mapsto \mathbf{c}(z) := c(z') \cap c(z'').$$

In a similar way, by Lemma 2.3 we can define the analogous conductor map $\mathbf{c}'$ for G'.

Lemma 2.7. *Suppose that the conductor c is coprime to both v and $\mathfrak{n}$, and that v is coprime to $\mathfrak{n}$. Let s be a geometric point in the special fiber at v, which lies in the supersingular locus if B_v is unramified and in the superspecial locus if B_v is ramified. Accordingly, let R_s as in Section 2.3.4. Then we have*

$$\mu^{\star}_{D,c}(s) = \frac{h(\mathscr{O}_{D,c}, R_s)}{2^{\omega(\mathfrak{n})+\varepsilon_\star} \# \operatorname{Pic}(\mathscr{O}_{D,c})},$$

for $\star \in \{ss, ssp\}$ and $\varepsilon_{ss} = 1$ and 0 otherwise.

Proof. We recall that $h(\mathscr{O}_{D,c}, R_s)$ denotes the number of equivalence classes modulo $R_s^\times$ of optimal embeddings $\mathscr{O}_{D,c} \hookrightarrow R_s$. By Lemma 2.1 we have that there are $2^{\omega(\mathfrak{n})}$ Galois orbits of CM points of conductor c and discriminant D, and that each such orbit has cardinality $\# \operatorname{Pic}(\mathscr{O}_{D,c})$.

By [8, Cor.3.2], we have that CM points of conductor c and discriminant D reducing to s are in bijection with the equivalence classes of the aforementioned optimal embeddings. In particular we have

$$2^{\varepsilon_\star} \cdot \#(\mathbf{c}^{-1}(c) \cap red_v^{-1}(s)) = \#(\mathbf{c}'^{-1}(c) \cap \pi^{-1}(s)) \tag{2.6}$$

where in the superspecial case $\varepsilon_{\mathrm{ssp}} = 0$ by the ramification of v in K following [8, Sec.3.4]. By Lemma 2.1, we obtain

$$h(\mathscr{O}_{D,c}, R_s) = 2^{\omega(\mathfrak{n})+\varepsilon_\star} \cdot \#(red^{-1}(s) \cap \Gamma_{D,c}).$$

Since the summation in (2.5) gives $\#\{z \in \Gamma_{D,c} : red_v(z) = s\}$, this shows that

$$\mu^{\star}_{D,c}(s) = \frac{\#\{z \in \Gamma_{D,c} : red_v(z) = s\}}{\#\Gamma_{D,c}} = \frac{h(\mathscr{O}_{D,c})}{2^{\omega(\mathfrak{n})+\varepsilon_\star} \# \operatorname{Pic}(\mathscr{O}_{D,c})}.$$

□

Remark 2.8. In [8, Thm.3.1], the lifting θ_v does not induce a bijection, but rather a κ-to-1 surjection, where κ has the cardinality of the Galois group $\operatorname{Gal}(K[c]/K[c_v])$ of the ring class field $K[c]$ as a factor, where c_v is the prime-to-v part of c. Therefore, our hypothesis that c is coprime to v implies the triviality of such a Galois group. The remaining factor of κ, which grosso modo is the cardinality of the set of $K_v^\times$-orbits of pairs of vertices of the Bruhat-Tits tree of $\mathrm{PGL}_2(F_v)$, is again 1 by our coprimality hypothesis combined with [8, Lemma 2.1(i)].

3. Equidistribution and Quadratic Forms

3.1. Automorphic Forms and Representations of Half-Integral Weight.

3.1.1. *Hilbert Modular Forms of Half-Integral Weight.* In this section we begin with some basic notions on Hilbert modular forms of weight $k = (k_1, ..., k_d) \in \frac{1}{2}\mathbb{Z}^d_{\geqslant 0}$. We set $\operatorname{Hom}(F, \mathbb{R}) = \{\tau_1, \dots, \tau_d\}$ and denote $z = (z_1, \dots, z_d) \in \mathscr{H}^d$. We invite the reader to go through [43] as a basic reference.

Remark 3.1. In what follows the reductive group GL_2 has nothing to do with the quaternionic setting of the Shimura curve (2.2), since the equidistribution result we aim to makes an auxiliary use of these automorphic forms.

Let k be as above and let $m \in \mathbb{Z}^d_{\geqslant 0}$. Let also U be a compact open subgroup of $\mathrm{GL}_2(\mathbb{A}_f)$. A complex *Hilbert modular form* of weight k, level U, character ω, is a (non necessarily holomorphic) function

$$f\colon \mathrm{GL}_2(\mathbb{A}_F) \to \mathbb{C}$$

such that:

(1) for all $g \in \mathrm{GL}_2(\mathbb{A})$, $\gamma \in \mathrm{GL}_2(F)$, $h \in U$,

$$f(\gamma g h) = J_k(h_\infty, z) f(g)$$

where J_k is the automorphy factor defined combining [43, 3.1b p.777] and [43, 1.4 p.770] and h_∞ is the archimedean part of h;

(2) there is a Whittaker–Fourier expansion

$$f(g) = C_f(g) + \sum_{\delta \in F^\times} W_f\left(\begin{bmatrix}\delta & \\ & 1\end{bmatrix} g\right) \tag{3.1}$$

where $C_f(g)$ is a constant term, while

$$W_f(g) = \int_{\mathbb{A}/F} f\left(\begin{bmatrix}1 & x\\ & 1\end{bmatrix} g\right)\omega(-x)dx$$

where dx denotes the obvious Haar measure induced by the adelic one.

We say that f is *cuspidal* if $C_f(g)$ is identically zero.

Let us denote by $e(x)$ the standard additive character of $\mathbb{A}_F$, i.e., $e(x) = \prod_v e_v(x_v)$ where $e_v(x_v) = \exp(2\pi i\{tr_{F_v/\mathbb{Q}_p}(x_v)\}_p)$, where $\{a\}_p$ is the p-fractional part of $a \in \mathbb{Q}_p$. If f is holomorphic they vanish unless $y_\infty > 0$. If this is the case then the Whittaker functions have the following expression

$$W_f\left(\begin{bmatrix}y & x\\ & 1\end{bmatrix}\right) = \tilde{a}(f, y)e_\infty(iy_\infty)e(x)$$

involving a function $\tilde{a}(f, y)$ of $y \in \mathbb{A}_f^\times$ which are called *Whittaker–Fourier coefficients* of f. For a finite idele $y = (y_v)_v \in \mathbb{A}_{F,f}^\times$, then one classically obtain a fractional ideal

$$\mathfrak{y} = y\widehat{\mathscr{O}}_F \cap F = \prod_v \mathfrak{p}^{\mathrm{val}_v(y_v)}, \tag{3.2}$$

where $\mathfrak{p}$ corresponds to the finite place v of F. The Whittaker–Fourier coefficients of f can be rewritten in terms of a function $a(f, \mathfrak{y})$ on the fractional ideals of F vanishing on the non-integral ideas. For a more detailed discussion, see [48, p.71] and [38, Sec.4.1.2].

We denote by $M_k(U, \omega)$ the space of holomorphic Hilbert modular forms of level U and character ω, and by $S_k(U, \omega)$ its subspace of cuspidal forms.

Let now the map $\mathfrak{a}\colon \mathrm{GL}_1(\mathbb{A}_{F,f}) \to \mathrm{GL}_2(\mathbb{A}_{F,f})$ be defined by $a \mapsto \begin{bmatrix}a & \\ & 1\end{bmatrix}$. We consider, following [12, (2.1.3)], the q-expansion map on $M_k(U, \omega)$

$$f \mapsto (\delta \to \tilde{a}(f, \delta y)). \tag{3.3}$$

This map is injective, by the q-expansion principle (see [12, Prop.2.1.1]), for $\delta y \in \mathfrak{a}^{-1}(U)$.

3.1.2. *Metaplectic Covers.* In this section we introduce a 2-fold cover of GL_2, as in [22]. Consider the local metaplectic group $\widetilde{\mathrm{SL}}_2$ given by

$$\widetilde{\mathrm{SL}}_2(F_v) = \begin{cases} \mathrm{SL}_2(F_v) \times \mathbb{Z}/2\mathbb{Z}, & \text{for v archimedean} \\ \text{the non-split central extension of } \mathrm{SL}_2(F_v) \text{ by } \mathbb{Z}/2\mathbb{Z}, & \text{otherwise.} \end{cases}$$

which is a double cover of SL_2. Note that the central extension of $\mathrm{SL}_2(F_v)$ is determined by the non-trivial element of the continuous group cohomology $\mathrm{H}^2_{\mathrm{cont}}(\mathrm{SL}_2(F_v), \mathbb{Z}/2\mathbb{Z}) \simeq \mathbb{Z}/2\mathbb{Z}$.

The group

$$\left\{\begin{bmatrix}a & \\ & 1\end{bmatrix} : a \in F_v^\times\right\}$$

gives an action of $F_v^\times$ on $\mathrm{SL}_2(F_v)$ by conjugation, which uniquely lifts to an automorphism of $\widetilde{\mathrm{SL}}_2(F_v)$. We denote by $\widetilde{\mathrm{GL}}_2(F_v)$ the semi-direct product of $\widetilde{\mathrm{SL}}_2(F_v)$ and $F_v^\times$. Hence this locally compact group fits in the following short exact sequence

$$1 \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow \widetilde{\mathrm{GL}}_2(F_v) \longrightarrow \mathrm{GL}_2(F_v) \longrightarrow 1.$$

The center of $\widetilde{\mathrm{GL}}_2(F_v^\times)$ is $Z^2 \times \mathbb{Z}/2\mathbb{Z}$, where $Z = \left\{\begin{bmatrix}a & \\ & a\end{bmatrix} : a \in F_v^\times\right\}$.

Globally, the metaplectic group is defined as the quotient

$$\widetilde{\mathrm{GL}}_2(\mathbb{A}_F) = \prod_v{}' \widetilde{\mathrm{GL}}_2(\mathscr{O}_{F_v})/\widetilde{Z},$$

where $\widetilde{Z} = \{\prod_v \epsilon_v \in \prod_v \mathbb{Z}/2\mathbb{Z} \mid \epsilon_v = 0 \text{ for an even number of } v\}$.

We conclude this section by introducing a dictionary between certain Hilbert modular forms and automorphic representations of $\widetilde{\mathrm{GL}}_2(\mathbb{A})$, following [38] and [20].

We say a cusp form f is *primitive* is it is a newform, an eigenform for all Hecke operators as defined in [38, Sec.4.1.3], and normalized with conductor equal to 1.

Lemma 3.2. *Let ω be a character of $(\mathscr{O}_F/\mathfrak{n})^\times$ and let $\tilde{\omega}$ denote its adelization, i.e., a character of $\mathbb{A}_F^\times/F^\times$ induced from ω. Let also D_i be the discrete series representation of $\widetilde{\mathrm{GL}}_2(\mathbb{R})$ of lowest weight $k_i \in \frac{1}{2}\mathbb{Z}$ with central character $x \mapsto |a|^{k_i}$. We have a bijection*

$$\left\{ \begin{array}{c} \textit{primitive cusp Hilbert modular forms} \\ \textit{of weight } k = (k_1, \dots, k_d) \in \frac{1}{2}\mathbb{Z}^d, \\ \textit{level } \mathfrak{n}, \textit{ nebentypus } \omega \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \textit{cuspidal automorphic representations of } \widetilde{\mathrm{GL}}_2(\mathbb{A}_F) \\ \textit{of conductor } \mathfrak{n}, \textit{ central character } \tilde{\omega} \\ \textit{whose representation at infinity is } \otimes_{j=1}^n D_{k_j - 1} \end{array} \right\}.$$

Proof. In view of the metaplectic theory developed in [20], the proof consists of an adaptation of [38, Thm.1.4], so we just summarize the above correspondence. In particular, at the archimedean places, the result follows from [20, Sec.4.1] which consists of the metaplectic counterpart of Langlands' classification for $\mathrm{GL}_2(\mathbb{R})$.

Given a primitive Hilbert cusp form f of half-integral weight k, we consider the space H_f spanned by right $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$-translations of f. We thus obtain a representation $\Pi(f)$ on H_f which occurs in the regular representation on the cusp forms. The representation $\Pi(f)$ is irreducible; this can be proved verbatim as in [38, Thm.4.7].

On the other hand, let Π be a cuspidal automorphic representation as in the left-hand side of the bijection and let V_Π be its representation space. The Whittaker model of Π is isomorphic to V_Π and for specific choices of vectors in the local Whittaker model of Π_v one can determine uniquely an element f in V_Π giving the desired Hilbert cusp form. □

3.1.3. *Weil Representations.* Let ψ_v denote the local standard additive character, and $\psi = \prod_v \psi_v$ the adelic additive character on $\mathbb{A}_F/F$. For each v, let here $\delta_v \in F_v$ be the conductor of ψ_v. For v finite, $\delta_v \mathscr{O}_{F_v}$ is the maximal fractional ideal over which ψ_v is trivial. On the other hand, for v a real place, $\psi_v(x) = \exp(2\pi i \delta_v x)$. Moreover, we have that the absolute value of $\delta := \prod_v \delta_v \in \mathbb{A}_F^\times$ equals the discriminant of F; in symbols, $|\delta| = d_F$.

Let $\mathtt{S}(F_v^n)$ denote the space of $\mathbb{C}$-valued Schwartz-Bruhat functions. For $f \in \mathtt{S}(F_v^n)$, we define its Fourier transform as in [20, p.36] by

$$\widehat{f}(x) = \int_{F_v^n} f(y)\psi_v(2xy)dy$$

where dy is the normalized Haar measure so that $\widehat{\widehat{f}}(x) = f(-x)$.

Let Q be a quadratic form on F^n, and let γ_v be a eight root of unity, and $\gamma_{v,a}$ its translate by $a \in F_v^\times$. The *local* Weil representation $r(\psi_v)$ is the unique representation of $\widetilde{\mathrm{SL}}_2(F_v)$ that can be realized on $\mathtt{S}(F_v^n)$ by the following formulae:

- $r(\psi_v)\begin{bmatrix} & 1 \\ -1 & \end{bmatrix} f(x) = \gamma_v \widehat{f}(x)$;
- $r(\psi_v)\begin{bmatrix} 1 & t \\ & 1 \end{bmatrix} f(x) = \psi_v(tQ(x))f(x)$ for $t \in F_v$;

- $r(\psi_v)\begin{bmatrix} a & \\ & a^{-1}\end{bmatrix} f(x) = |a|_v^{1/2}\dfrac{\gamma_v}{\gamma_{v,a}} f(x)$;
- $r(\psi_v)(\varepsilon)f(x) = \varepsilon f(x)$ for ε is a second root of unity.

Note that these formulae uniquely define the Weil representation via the Bruhat decomposition of SL_2.

Globally, for a totally real field F and a quadratic space (V, Q) over F, consider the non-trivial character $\psi\colon \mathbb{A}_F/F \to \mathbb{C}^\times$. Then the Weil representation is obtained as the restricted tensor product of the local Weil representations $r(\psi_v)$'s.

Consider now the quadratic space (V, Q) over F_v, for Q a n-ary form. By [20, Ex.2.21] each quadratic form Q corresponds to a Weil representation. Whenever in need to emphasize this we will denote the Weil representation corresponding to Q by r_Q.

3.1.4. *Theta-representations.* We now focus on the unary case. The local Weil representation $r(\psi_v)$ decomposes as a direct sum of two irreducible subrepresentations on the subspace even and odd Schwartz-Bruhat functions. Thus, for χ_v character of $F_v^\times$, we can tensor with the even or odd part of $r(\psi_v)$, according to the parity of χ_v, and inducing this representation to $\widetilde{\mathrm{SL}}_2(F_v^n)$ we obtain an irreducible admissible representation $r(\chi_v)$ independent of ψ_v, which is unramified whenever χ_v is unramified, i.e., for all but finitely many v.

We now extend $r(\psi)$ to a representation of $\widetilde{\mathrm{GL}}_2(F_v^n)$, so to remove the dependence on ψ_v.

Consider the pullback of $\{g \in \mathrm{GL}_2(F_v^n) : \det(g) \in (F_v^\times)^2\}$ in $\widetilde{\mathrm{GL}}_2(F_v^n)$ taking the form

$$\widetilde{G}^* := \widetilde{\mathrm{SL}}_2(F_v) \rtimes \left\{\begin{bmatrix} 1 & \\ & a^2\end{bmatrix} : a \in F_v^\times\right\}.$$

We then extend $r(\psi_v, \chi_v)$ to $\widetilde{G}^*$ by setting

$$r(\psi_v, \chi_v)\begin{bmatrix} 1 & \\ & a^2\end{bmatrix} f(x) = \chi_v(a)|a|_v^{-1/2} f(a^{-1}x).$$

Moreover, we have that $r(\psi_v, \chi_v)$ is an irreducible, admissible representation of $\widetilde{G}^*$.

Inducing up $r(\psi_v, \chi_v)$ to $\widetilde{\mathrm{GL}}_2(F_v^n)$ produces a representation $r(\chi_v)$, which is irreducible and admissible and independent of ψ_v (see [21, 1.3, p.150]).

We conclude by the global setting. Recall that an automorphic representation of half-integral weight is an irreducible admissible representation of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$ which is isomorphic to a subrepresentation of the space of automorphic forms on $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$.

For every character χ of $\mathbb{A}^\times/F^\times$, the (global) theta-representation is

$$r(\chi) = \otimes_v' r(\chi_v),$$

namely the restricted tensor product of the local Weil representations. This gives an automorphic representation of half-integral weight. Moreover, if there exists at least a place v of F such that χ is odd (i.e., $\chi_v(-1) = -1$), then $r(\chi)$ is cuspidal (see [21, Prop.8.1.1]).

Let us introduced two important technical definitions. We say that an irreducible admissible representation $\overline{\pi}_v$ of $\widetilde{\mathrm{GL}}_2(F_v)$ is *v-distinguished* if its Whittaker model, which always exists, is unique. Globally, an irreducible admissible representation $\overline{\pi} = \otimes' \overline{\pi}_v$ of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$, is distinguished if it is v-distinguished at all places v. An automorphic representation of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$ is called *genuine* if it does not factor through $\mathrm{GL}_2(\mathbb{A}_F)$.

Lemma 3.3. *There is a bijection*

$$\left\{\ \text{Hecke characters of } F\ \right\} \longleftrightarrow \left\{\begin{array}{c}\text{genuine, distinguished, automorphic}\\ \text{representations of } \widetilde{\mathrm{GL}}_2(\mathbb{A}_F)\end{array}\right\}$$

attaching to every Hecke character χ the theta-representation $r(\chi)$.

Proof. See [21, Thm.A p.147]. $\square$

Remark 3.4. We will make extensive use of the fact that $r(\chi)$ generalizes the construction that associates to a Dirichlet character χ on $\mathbb{Z}/N\mathbb{Z}$ the classical theta series

$$\theta(z,\chi) = \sum_{n\geqslant 1} \chi(n) n q^{n^2 z}.$$

It is well known that this gives a primitive modular form of half-integral weight, level $4N^2$ and character χ. Moreover, for t squarefree we have that the subspace $\mathcal{U}(N,\chi)$ of cuspidal modular form of weight $3/2$ whose Shimura lift is *not* cuspidal is spanned by $\theta(z,\chi)$'s (see, for instance, [24, p.365]).

3.1.5. *Eisenstein Series.* Let B denote the parabolic group of GL_2, i.e., its Borel subgroup, with Levi decomposition $B = NM$ for N being the unipotent subgroup and M the Levi subgroup. Denote also by $\mathrm{GL}_2(\mathbb{A}_F)^1$ the element of $\mathrm{GL}_2(\mathbb{A}_F)$ whose determinant has norm 1, and by $M(\mathbb{A}_F)^1$ the matrices in the Levi subgroup whose entries are of norm 1.

For $\varphi \in \mathcal{C}_c^\infty(N(\mathbb{A}_F)M(\mathbb{A}_F)\backslash \mathrm{GL}_2(\mathbb{A}_F)^1)$ we define the *pseudo-Eisenstein* series as

$$\Psi(\varphi, g) = \sum_{\gamma\in B(F)\backslash \mathrm{GL}_2(F)} \varphi(\gamma g).$$

Note that $\Psi(\varphi,g)$ is locally finite[11], so it converges and defines and element of $\mathcal{C}_c^\infty[\mathrm{GL}_2(\mathbb{A}_F)^1]$ (see [19, 2.7.1]).

Let now χ be a character of $M(F)\backslash M(\mathbb{A}_F)^1$. Note that any character of $[M(\mathbb{A}_F)^1]$ is of the form $\nu^s\chi$ where $\nu^s\colon \begin{bmatrix}\delta(y) & \\ & 1\end{bmatrix} \mapsto |y|^s$ for $\delta\colon (0,\infty)\to \mathbb{A}_F^\times$ is the diagonal embedding into the archimedean component of the ideles by $\delta(y) = (\dots, y^{\frac{1}{d_v}}, \dots)$ with d_v the local degree at v. We thus consider the following unramified principal series representation of GL_2 given by parabolic induction

$$I_{s,\chi} = \{f \in \mathcal{C}_c^\infty(N(\mathbb{A}_F)M(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)^1) : f(nmg) = (\nu^s\chi)(m)f(g) \text{ for all } n\in N(\mathbb{A}_F), m\in M(\mathbb{A}_F)\}.$$

The *Eisenstein* series is defined as

$$E(f,g) = \sum_{\gamma\in B(F)\backslash \mathrm{GL}_2(F)} f(\gamma g),$$

for $f \in I_{s,\chi}$. Note that the Eisenstein series is not a $\mathscr{L}^2$-function[12], while the pseudo-Eisenstein series is. Moreover, the pseudo-Eisenstein series can be expressed in terms of Eisenstein series by integral converging uniformly absolutely on compacts in $\mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)^1$. For the pointwise formula, see [19, Thm.2.11.1].

3.1.6. *Automorphic Shimura Lifting.* In the local case, we have that an irreducible admissible representation π_v of $\mathrm{GL}_2(F_v)$ is the local Shimura image of $\overline{\pi}_v$, in symbols, $\mathrm{Shi}(\overline{\pi}_v) = \pi_v$ if

- for $a \in F_v^\times$, we have
$$\omega_v(a) = \overline{\omega}_v(a^2),$$
where $\omega_v, \overline{\omega}_v$ are the central characters of π_v and $\overline{\pi}_v$ respectively;
- for any character χ of $F_v^\times$, equalities between automorphic L-functions and ε-factors hold respectively as in [22, Sec.7.1].

[11] I.e., for g in a fixed compact of $\mathrm{GL}_2(\mathbb{A}_F)$, the sum defining $\Psi(\phi,g)$ has only finitely many summands.
[12] The obstruction to this consists of the constant term.

The Shimura lift of $\overline{\pi}$, if it exists, is unique[13]. For a L-functions-free approach to the Shimura lifting, see [16, Sec.5].

As usual, we define the global Shimura lifting by putting the local pieces together. Let $\overline{\pi} = \otimes_v' \overline{\pi}_v$ be an irreducible admissible representations of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$. We say that the irreducible admissible representation π of $\mathrm{GL}_2(\mathbb{A}_F)$ is the *Shimura image* of $\overline{\pi}$, in symbols, $\mathrm{Shi}(\overline{\pi}) = \pi$, if, for every place v, the same holds locally, i.e., $\mathrm{Shi}(\overline{\pi}_v) = \pi_v$.

Let now $\chi = \prod_v \chi_v$ be a Hecke character of F, and consider the case $\overline{\pi} = r(\chi)$. By [22, Thm.15.1] we obtain that $\mathrm{Shi}(r(\chi_v))$ is 1-dimensional and unramified for almost every place v, and so

$$\mathrm{Shi}(r(\chi)) = \otimes_v' \mathrm{Shi}(r(\chi_v)) \tag{3.4}$$

is automorphic but not cuspidal. In particular, $\mathrm{Shi}(\overline{\pi})$ is cuspidal if and only if $\overline{\pi}$ is not in the image of the correspondence of Lemma 3.3.

In the archimedean case F_∞, then $\overline{\pi}_\infty$ is a discrete series representation of lowest weight $k/2$ and so $\mathrm{Shi}(\overline{\pi}_\infty)$ corresponds to a discrete series representation of lowest weight $k-1$ (see [20, Prop.4.8]).

Lastly, we extend the Serre–Stark theorem (and the considerations of Remark (3.4)) to this automorphic setting as in [21]. This means that the theta representations $r(\chi)$'s exhaust certain automorphic representations of weight 1/2, which are defined to be such that there exists one archimedean place such that $\overline{\pi}_\infty$ is the even part of the Weil representation.

Lemma 3.5. *Suppose that $\overline{\pi}$ is a cuspidal representation of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$ of weight 1/2. Then there exists a Hecke character χ of F such that $\overline{\pi} = r(\chi)$.*

Proof. This follows as a consequence of the surjective part of Lemma 3.3 and the Shimura lift. □

3.1.7. *Langlands Spectral Decomposition.* In this section we mainly follow [2] and [34] as references.

Let $\mathscr{L}^2[\mathrm{GL}_2]$ denote the Hilbert space of functions on $[\mathrm{GL}_2]$ that are square-integrable with respect to the natural measure, and denote by R the right regular representation of $\mathrm{GL}_2(\mathbb{A}_F)$ on $\mathscr{L}^2[\mathrm{GL}_2]$. This representation is isomorphic to the completion of a discrete sum of unitary irreducible representations of $\mathrm{GL}_2(\mathbb{A}_F)$.

Consider $\mathrm{GL}_2(\mathbb{A}_F)^1 = \{g \in \mathrm{GL}_2(\mathbb{A}_F) : |\det g|_{\mathbb{A}} = 1\}$ where $|\cdot|_{\mathbb{A}}$ is the adelic norm. Denote by $\mathscr{L}$ the space $\mathscr{L}^2(\mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)^1))$. We want to introduce the spectral decomposition of $\mathscr{L}$ under R.

Let us introduce the *discrete* spectrum of $\mathscr{L}$ which we call $\mathscr{L}_{\mathrm{disc}}$. It is defined as a Hilbert sum of irreducible subspaces of $\mathscr{L}$, i.e., the closed subspace generated by the irreducible subrepresentations of $\mathrm{GL}_2(\mathbb{A}_F)^1$ in $\mathscr{L}$. Thus $\mathscr{L}_{\mathrm{disc}}$ is the part of $\mathscr{L}^2[\mathrm{GL}_2(\mathbb{A}_F)^1]$ whose spectral decomposition looks like the decomposition of R. The complement of $\mathscr{L}_{\mathrm{disc}}$ in $\mathscr{L}$ is called the *continuous* spectrum, and denoted by $\mathscr{L}_{\mathrm{cont}}$.

Let B be the Borel subgroup of GL_2 and N its unipotent subgroup. Consider the square-integrable function on $[N] = N(\mathbb{A}_F) \cap \mathrm{GL}_2(F)\backslash N(\mathbb{A}_F)$ defined for almost all $x \in \mathrm{GL}_2(\mathbb{A}_F)^1$ as

$$n \mapsto \varphi(nx)$$

for $\varphi \in \mathscr{L}$.

We define the *constant term* along B the function φ_B defined as

$$\varphi_B(x) = \int_{[N]} \varphi(nx)dx,$$

where dx is the Haar measure on $[N] \simeq F\backslash\mathbb{A}_F$. The function φ is said to be *cuspidal* is its constant term along B vanishes almost everywhere. We thus define $\mathscr{L}_{\mathrm{cusp}}$ to be the subspace of cuspidal functions

[13]Moreover, as suggested by its notation, it does not depend on ψ_v, while the ε-factor does.

$\varphi \in \mathscr{L}$. It can be decomposed as the Hilbert sum of unitary irreducible representations of $\mathrm{GL}_2(\mathbb{A}_F)^1$ with finite multiplicities[14].

Next Lemma yields a construction of the orthogonal complement of $\mathscr{L}_{\mathrm{cusp}}$ in $\mathscr{L}$.

Lemma 3.6. *The $\mathscr{L}^2$-closure of the set of pseudo-Eisenstein series spans the orthogonal complement of $\mathscr{L}_{cusp}$.*

Proof. See [19, 2.7.2]. □

The orthogonal complement of $\mathscr{L}_{\mathrm{cusp}}$ in $\mathscr{L}_{\mathrm{disc}}$ is called the *residual* spectrum, and denoted by $\mathscr{L}_{\mathrm{res}}$. It is obtained by residues of Eisenstein series. In our situation we have that $\mathscr{L}_{\mathrm{res}}$ decomposes as the direct sum of $\chi \circ \det$, for χ ranging among the characters of $\mathbb{A}_F^1/F$.

Proposition 3.7. *We have the orthogonal decomposition*

$$\mathscr{L} = \mathscr{L}_{cusp} \oplus \mathscr{L}_{res} \oplus \mathscr{L}_{cont}. \tag{3.5}$$

Proof. This statement follows from [34, Thms 2.10.3, 2.10.4] and [34, Prop. 2.5.10]. □

3.2. **Ternary Theta Series.**

3.2.1. *Ternary Quadratic Forms.* Let V denote a 3-dimensional vector space over F. An $\mathscr{O}_F$-valued quadratic form Q over $\mathscr{O}_F$ is *primitive* if the ideal generated by its $\mathscr{O}_F$-values is $\mathscr{O}_F$. Moreover Q *represents* δ if $\delta \in Q(\mathscr{O}_F)$. We also denote the orthogonal group of a quadratic space (V, Q) by $O(V)$, which consists of the invertible linear transformations $L \in \mathrm{End}(V)$ such that $Q(Lv) = Q(v)$ for all $v \in V$. Moreover, rotations form the special orthogonal group $\mathrm{SO}(V)$, i.e., isometries of determinant 1. We also recall that a non-zero vector v is called *anisotropic* if $Q(v) \neq 0$. We also say that Q is *positive definite* if Q represents only positive values. For such a form Q, essential to our study are the *representation* numbers

$$r(Q, \delta) := \#\{x \in \mathscr{O}_F^3 : Q(x) = \delta\}.$$

Note that assuming the positive definiteness is essential, since it ensures the finiteness of the representation numbers.

We can now define the theta series of Q and its Fourier expansion. For $\delta \in F$, we denote by $q^\delta := \exp(2\pi i \sum_{i=1}^d \tau_i(\delta) z_i)$.

Let Q be a non-degenerate[15] positive definite ternary quadratic form on F^3. Then

$$\theta_Q(z) = \sum_{v \in \mathscr{O}_F^3} q^{Q(v)} = \sum_{\delta \in F} r(Q, \delta) q^\delta \tag{3.6}$$

is a Hilbert modular form of weight $3/2$ for $\mathrm{SL}_2(\mathscr{O}_F)$. See [19, p.154] or [43, Sec.4] for a proof.

From now on, let B^0 denote the set of trace zero element in the quaternion algebra B. To B, one can associate the 4-dimensional quadratic space $(V, Q) = (B, nr)$.

As already hinted in Section 3.1.3, we make clear that the cases of interest for the present work are the following:

- $(V, Q) = (B^0, Q_s)$;
- $(V, Q) = (F, U)$;

where U is the unary form $U(x) = x^2$ and Q_s is the ternary form as defined in Section 3.2.2.

---

[14] In fact, its multiplicities are either 0 or 1.

[15] I.e., $\det Q \neq 0$.

3.2.2. *Ternary Quadratic Forms and Eichler Orders.* The constructions of the Eichler orders attached to geometric points s in the special fiber of the Shimura curve as in Section 2.3.4 provide the *Gross lattice*

$$\Lambda_s := (2R_s + \mathscr{O}_F) \cap B'^0$$

and the ternary quadratic forms

$$Q_s \colon (2R_s + \mathscr{O}_F) \cap B'^0 \longrightarrow F,$$

defined by $Q_s(b) = nr(b)$.
Notice that, since B' is ramified at τ_1, then Q_s is positive definite.

Let us now denote by Λ_s the domain of Q_s. Extending the scalars by $\mathbb{A}_{F,f}$, we obtain the adelization of Q_s

$$\widehat{Q}_s \colon \Lambda_s \otimes \mathbb{A}_{F,f} \longrightarrow \mathbb{A}_{F,f}.$$

For $\delta \in \mathbb{A}_{F,f}^{\times}$, we consider the following representation numbers

$$r(Q_s, \delta) = \#\{x \in \Lambda_s \otimes \mathbb{A}_{F,f} : \widehat{Q}_s(x) = \delta \bmod \widehat{\mathscr{O}}_F^{\times}\}.$$

Let (V, Q) denote a quadratic space over F with a symmetric, bilinear form β, and let $(e_i)_{i=1}^d$ be a F-basis for V, and consider the $d \times d$-matrix $M = (\beta(e_i, e_j))_{i,j=1}^d$. Then the discriminant of the quadratic form Q over F is defined as the coset $D_Q = \det M \cdot F^{\times,2}$ in $F^\times / F^{\times,2}$.

Lemma 3.8. *Suppose that* 2 *is inert in* F. *Let* $\mathfrak{n}$ *be the level of* R_s. *The discriminant* D_{Q_s} *of the quadratic lattice* Λ_s *is equal to the* $4\mathfrak{n}^2 v^2$.

Proof. Let us begin with the case of a place v of F such that $v \neq \ell$. Let ϖ_v be the uniformizer of $\mathscr{O}_{F_v}$. Denote by n the v-adic valuation of $\mathfrak{n}$,so that $R_{s,v} := R_s \otimes \mathscr{O}_{F_v}$ is a Eichler order of level ϖ_v^n. We thus have

$$R_{s,v} = \begin{bmatrix} \mathscr{O}_{F_v} & \mathscr{O}_{F_v} \\ \varpi_v^n \mathscr{O}_{F_v} & \mathscr{O}_{F_v} \end{bmatrix}$$

and consequently

$$\Lambda_{s,v} := \Lambda_s \otimes \mathscr{O}_{F_v} = \left\{ \begin{bmatrix} a & 2b \\ 2\varpi_v^n c & -a \end{bmatrix} : a, b, c \in \mathscr{O}_{F_v} \right\}.$$

Hence, by computing the determinant, the local quadratic form $Q_{s,v}$ is

$$(a, b, c) \mapsto -a^2 - 4\varpi_v^{2n} bc.$$

This shows that the contributions to the determinant of Q_s are

$$\begin{cases} \varpi_v^{2n}, & \text{if } v \not| 2\mathscr{O}_F \\ 4\varpi_v^{2n}, & \text{otherwise.} \end{cases}$$

Consider now the case $v = \ell$. If $v \neq 2\mathscr{O}_F$, then $R_{s,v}$ is the unique maximal order of the unique quaternion algebra, with $\mathscr{O}_{F_v}$-basis $(1, i, j, k)$ such that $i^2 = -\varpi_v$, $j^2 = -1$, $k^2 = -1$ and $k = ij = -ji$. Thus $\Lambda_{s,v}$ has a basis $(2i, 2j, 2k)$ amd we obtain the diagonal form

$$Q_{s,v}(a, b, c) = 4\varpi_v a^2 + 4b^2 + 4\varpi_v c^2$$

whose determinant is $64\varpi_v^2$, so contributing by ϖ_v^2 to the discriminant of Q_s.
On the other hand, if $v = 2\mathscr{O}_F$, then by [23, pp.145,177], the local Eichler order is isomorphic to the unique maximal order and for $i^2 = j^2 = k^2 = -1$ and $k = ij = -ji$ we have

$$\Lambda_{s,v} = \{ai + bj + ck : a \equiv b \equiv c \bmod v\}.$$

Therefore we have

$$Q_{s,v}(a, b, c) = -(3a^2 + 4ab + 4ac + 4b^2 + 4c^2),$$

whose corresponding matrix has determinant 16, i.e., it contributes $4\ell^2$. ∎

Let $r(Q_s,\delta)$ be the number of representations of $\delta \in \mathbb{A}_F^\times$ by $\widehat{Q}_s$. We introduce, following [26], the Kohnen plus space $M^+_{\frac{3}{2}}(U)$ as the subspace of $M_k(U)$ whose forms f have Fourier coefficients $a(f,\delta)=0$, for $\delta \in \mathbb{A}_F^\times$, unless δ_v is congruent to a square modulo $4\mathscr{O}_{F_v}$ for almost all v, i.e., there exists $y_v \in \mathscr{O}_{F_v}$ such that $\delta_v \equiv y_v^2 \bmod 4\mathscr{O}_{F_v}$ for almost all v.

Lemma 3.9. *The Gross theta series*

$$\theta_{Q_s} = \sum_{\beta\in\Lambda_s} q^{Q_s(\beta)} \tag{3.7}$$

are Hilbert modular forms which lie in the Kohnen's plus space $M^+_{\frac{3}{2}}(\Gamma_0(4\mathfrak{n}\ell))$.

Proof. By the q-expansion map (3.3) we have that the theta series θ_{Q_s} are Hilbert modular forms and by (3.6) it follows than its weight is $3/2$.

For $\delta_v \in \mathscr{O}_{F_v}$, we notice that $\mathscr{O}_{F_v} + \dfrac{\delta_v+\sqrt{-\delta_v}}{2}\mathscr{O}_{F_v}$ is an $\mathscr{O}_{K_v}$-order if and only if $\delta_v \equiv 0,1 \bmod 4\mathscr{O}_{F_v}$.

Let us show the condition under which it is an $\mathscr{O}_{K_v}$-order. We require that $(\delta_v+\sqrt{-\delta_v}/2)^2 \in \mathscr{O}_{F_v}$, which is equivalent to $(\delta_v^2-\delta_v)/4 \in \mathscr{O}_{F_v}$. Hence the sufficient and necessary condition to be an order is that $\delta_v \equiv 0,1 \bmod 4\mathscr{O}_{F_v}$. Since $Q_{s,v}(\beta_v) = -\delta_v \equiv 0,1 \bmod 4\mathscr{O}_{F_v}$ we have that the coefficients of $\theta_{Q_{s,v}}$, i.e., the $r(Q_{s,v},\delta_v)$'s are non-zero only if $-\delta_v \equiv 0,1 \bmod 4\mathscr{O}_{F_v}$. By Hasse–Minkowski theorem we obtain the global result.
Concerning the level, we proceed as in [27, Lemma 4.4]. Given that that $\theta_{Q_s} \in M^+_{3/2}(\Gamma_0(4\mathfrak{m}))$, for $\mathfrak{m}$ the smallest ideal such that $4\mathfrak{m}$ has even integral coefficients, it is enough to see that the inverse of the matrices appearing in the proof of Lemma (3.8) have even integral coefficients after multiplying by $4v^n$, where n is the v-adic valuation of $\mathfrak{n}\ell$. ☐

3.2.3. *Automorphic Theta Series.* In order to obtain automorphic theta forms from Weil representations, we need to introduce the analogue of self-dual[16] functions on the lattice $V(F) \simeq F^n$, for a quadratic space (V,Q) with Q a n-ary form.

For v a non-archimedean place of F, consider the functions

$$\mathbf{1}_v := \prod_{i=1}^n \mathbf{1}_{\mathscr{O}_{F_v}},$$

where $\mathbf{1}_{\mathscr{O}_{F_v}}$ denotes the characteristic function of $\mathscr{O}_{F_v}$. We have that $\widehat{\mathbf{1}}_v = \mathbf{1}_v$, i.e., the Fourier transform of $\mathbf{1}_v$ is itself. For $v=\infty$ an archimedean place, we define $\mathbf{1}_\infty$ to be the Gaussian exponential $\exp(-\pi Q(x))$. Note that $\mathbf{1}_v \in \mathscr{L}^2(F_v^n)$. Any adelic Schwartz-Bruhat function $\phi \in \mathtt{S}(\mathbb{A}_F^n)$ is a (finite) linear combination of the product $\prod_v \phi_v$ where $\phi_v \in \mathtt{S}(F_v^n)$ and $\phi_v = \mathbf{1}_v$ for almost all v (see [25, Rmk.4.3.1]).

In order to associate to each of these quadratic forms a function on $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$ we consider the following linear functional on $\mathtt{S}(\mathbb{A}_F^n)$

$$\Theta\colon \phi \mapsto \sum_{x\in F^n} \phi(x)$$

called *theta*-distribution. Note that the series defining the $\Theta(\phi)$ converges and Θ spans the space of $\mathrm{SL}_2(F)$-invariant forms on $\mathtt{S}(\mathbb{A}_F^n)$. After inducing up to $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$, the action of the Weil representation r_Q attached to the form Q on the theta-distribution defines

$$\vartheta_Q(g,\phi) = \sum_{v\in\mathbb{A}_F^n} r_Q(g)\phi(v)$$

for $g \in \widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$. This function is called *theta-kernel* and it defines an automorphic form on $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$. For more details, see [20, Sec.2.6].

[16] With respect to the Fourier transform.

3.2.4. *Ternary Quadratic Forms and Optimal Embeddings.* In this section we introduce one of the key ingredient of the equidistribution we are going to prove.

Lemma 3.10. *Let δ_c be the idele corresponding to the ideal Dc^2 by (3.2). Then we have the following $((\#R_s^\times/\#\mathscr{O}_{D,c}^\times):1)$-correspondence*

$$\left\{ \begin{array}{c} \textit{primitive representations} \\ \widehat{Q}_s(b) = \delta_c \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \textit{optimal embeddings} \\ f\colon \mathscr{O}_{D,c} \hookrightarrow R_s \end{array} \right\}.$$

Proof. First we show how to obtain a representation from an embedding and viceversa, inspired by [23, Prop.12.9]. Indeed, we proceed locally at the finite places v of F. Let $\delta_{c,v} = D_v c_v$ be the product of the generators of the ideals D and c at v. By the local-global principle for embeddings (see [45, Prop.14.6.7]), we can write

$$\mathscr{O}_{D,c,v} = \mathscr{O}_{F_v} + \frac{\delta_{c,v} + \sqrt{-\delta_{c,v}}}{2}.$$

Consider an embedding $f_v\colon \mathscr{O}_{D,c,v} \hookrightarrow R_{s,v}$ and denote by $\beta_v := f_v(\sqrt{-\delta_{c,v}})$. Since

$$2f_v\left(\frac{1}{2}(\delta_{c,v} + \sqrt{-\delta_{c,v}})\right) = \delta_{c,v} + \beta_v,$$

then $\beta_v \equiv -\delta_{c,v} \bmod 2R_{s,v}$, so that $\beta_v \in \Lambda_{s,v}$. Since $nr_v(\beta_v) = \delta_v$, we obtained the desired representation.

On the other hand, let $\beta_v \in \Lambda_{s,v}$ of reduced norm $\delta_{c,v}$. We thus write $\beta_v = \gamma_v + 2r_v$, for $\gamma_v \in \mathscr{O}_{F_v}$ and $r_v \in R_{s,v}$, so that

$$\delta_{c,v} = nr_v(b_v) = -(\gamma_v + 2r_v)^2 \equiv -\gamma_v \bmod 4R_{s,v}.$$

Thus we have $\beta_v \equiv -\delta_{c,v} \bmod 2R_{s,v}$, and we obtain an embedding $f_v\colon \mathscr{O}_{D,c,v} \hookrightarrow R_{s,v}$ by setting

$$f_v\left(\frac{\delta_{c,v} + \sqrt{-\delta_{c,v}}}{2}\right) = \frac{\delta_{c,v} + \beta_v}{2}.$$

Finally, we recall that being optimal is a local property, i.e., an embedding $f\colon \mathscr{O}_{D,c} \hookrightarrow R_s$ is optimal if and only if $f_v\colon \mathscr{O}_{D,c,v} \hookrightarrow R_{s,v}$ is optimal at all the places v (see, for instance, [45, Lemma 30.3.6]). Therefore, we proceed by localizing as above.

Since we closely follow [27, Lemma 4.1], we only sketch the strategy to show the correspondence between optimal embeddings and primitive representations without the necessary algebraic manipulations. By contrapositive, if $Q_s(\beta_v) = \delta_{c,v}$ is not primitive, then we can write $\beta_v = \alpha_v k_v$ for $\alpha_v \in \Lambda_{s,v}$ and $k_v \in \mathscr{O}_{F_v}$. By showing that $\frac{\delta_{c,v} + \alpha k_v^2}{2k_v^2}$ belongs to $f_v(K_v) \cap R_{s,v}$ but not to $f_v(\mathscr{O}_{D,c,v})$, one concludes that f_v is not optimal. Lastly, if we suppose that f_v is not an optimal embedding, then $f_v(\mathscr{O}_{D,c,v} \otimes F_v) \cap R_{s,v} = \mathscr{O}_{D,c',v}$ where $c = c'k$ for $k \in \mathscr{O}_{F_v}$. □

3.2.5. *Genus and Spinor Genus.* We revise here a few classical concepts following [25]. An equivalent formulation can be found, for instance, in [28, p.105].

Let (V, Q) denote a quadratic space over F with a symmetric, bilinear form β. We recall that another form Q' is *equivalent* to Q if there exists an invertible linear change of variables f such that $Q'(f(x)) = Q(x)$. The set of $\mathscr{O}_F$-equivalence classes that are $\mathscr{O}_{F_v}$-equivalent locally for all v's is the *class* number of Q. The *genus* of Q is the set of all forms with the same localization as Q. The *class* number of Q is equal to the cardinality of gen(Q), which is finite by a classical result of Siegel. In terms of lattices, let L be a integral lattice in V. Then we denote by $gen(L)$ the genus of L, defined to be the set of integral lattices M such that $L_v \simeq M_v$ for every place v of F (including the archimedean ones). More geometrically, the genus of L is the orbit of L under the action of the orthogonal group. Note that lattices in the same genus are isomorphic as $\mathscr{O}_F$-modules.

In the case of ternary quadratic forms, the genus is moreover subdivived into *spinor* genera, which essentially are the subgenera given by equivalence under the spinor group. More precisely, the *spinor* norm is the group homomorphism

$$\begin{aligned} O(V) &\to F^\times/F^{\times,2} \\ (x \mapsto x - 2u\beta(x,u)/Q(u)) &\mapsto Q(u) \end{aligned}$$

i.e., it sends a reflection orthogonal to v into $Q(v)$. We denote by $\mathrm{Spin}(V)$ the isometries of spinor norm 1. This gives a two-fold cover of the special orthogonal group.
Two lattices L and M are in the same *spinor genus* if there is a rotation $\sigma \in \mathrm{SO}(V)$ and for every place v there is $r_v \in \mathrm{Spin}(V_v)$ such that $L_v \simeq \sigma_v r_v M_v$.

Lastly, we recall that the *automorphs* of Q are the isometries from Q to itself. Let w_Q be the number of the automorphs of Q, i.e., the cardinality of $\mathrm{Stab}_{\mathrm{GL}_3(\mathscr{O}_F)}(Q)$. Indeed the automorphs of Q are finite: since any of them is determined by its action on a $\mathscr{O}_F^3$-basis, and for m big enough we have that the finitely many vectors in the compact defined by $Q(u) \leqslant m$ generate $\mathscr{O}_F^3$.

3.2.6. *Theta Series associated to Genus and Spinor Genus.* We now associate two theta series to the genus $\mathrm{gen}(Q_s)$ and spinor genus $\mathrm{spn}(Q_s)$.
Let us define the genus and spinor genus *mass*[17]

$$r^?(\diamondsuit, \delta) = \frac{\sum_{Q\in\diamondsuit} r^?(Q,\delta)/w_Q}{\sum_{Q\in\diamondsuit} 1/w_Q} \tag{3.8}$$

and

$$\theta_\diamondsuit = \sum_{\delta\in\mathbb{A}_F} r(\diamondsuit,\delta) q^\delta \tag{3.9}$$

for $\diamondsuit \in \{\mathrm{gen}(Q_s), \mathrm{spn}(Q_s)\}$ and $? \in \{\emptyset, *\}$, where $*$ denotes the restriction to primitive representations[18].

Proposition 3.11. *Let s be in the supersingular or superspecial locus of the special fiber of $\mathscr{S}_U$ at v, for v unramified or ramified in K respectively. Then we have:*
1. *the theta series $\theta_\diamondsuit$'s are in the same space as θ_{Q_s};*
2. *the modular form*
$$\theta_{Q_s} - \theta_{\mathrm{spn}(Q_s)}$$
of weight $3/2$ lies in the orthogonal[19] *complement of the space of 1-dimensional theta series;*
3. *the modular form*
$$\theta_{\mathrm{gen}(Q_s)} - \theta_{\mathrm{spn}(Q_s)}$$
of weight $3/2$ lies in the space spanned by 1-dimensional theta series.

Proof. Let us first recall the very well known fact that the space of Hilbert modular forms decomposes uniquely as a direct sum of cusp forms and Eisenstein series. In particular, the Hilbert modular form θ_{Q_s} defined in Lemma 3.9 decomposes as

$$\theta_{Q_s} = E + H + s$$

namely, the sum of an Eisenstein series E and two cuspidal forms H and s, where the Shimura lift of H is an Eisenstein series, hence it is non-cuspidal.
In the automorphic setting, we have the automorphic theta form ϑ_{Q_s} and, by Lemma 3.2, H and s corresponds to two cuspidal automorphic forms which we still denote by π_H and π_s. By Lemma 3.16 and Proposition 3.7 we can decompose the automorphic theta form as

$$\vartheta_{Q_s} = \Psi + \pi_H + \pi_s$$

---

[17] From the German word "maß", which means "weight, measure".
[18] And $\emptyset$ symbolizes the absence of any index.
[19] Where orthogonality is considered with respect to the Petersson inner product.

for a pseudo-Eisenstein series Ψ and for $\pi_H \in \mathscr{L}_{\mathrm{res}}$. By Section 3.1.6, we have that the "kernel" of the Shimura lifting Shi consists of those cuspidal $\overline{\pi}$'s which come from Hecke characters. This immediately implies that π_H has a non-cuspidal Shimura lift.

Consider now the Whittaker–Fourier coefficients W_E and W_H of E and H. By the work of Schulze-Pillot [42, Satz 2] (see also [24, p.366])

$$W_E(\delta) = r(\mathrm{gen}(Q_s), \delta), \quad W_E(\delta) + W_H(\delta) = r(\mathrm{spn}(Q_s), \delta). \tag{3.10}$$

By (3.10) and the fact that the θ_Q's are all in the same Kohnen space as θ_{Q_s} for $Q \in \mathrm{gen}(Q_s)$ we obtain that $\theta_\diamond$'s are in the same space as θ_{Q_s}.

We now prove the second part of the Lemma. In the automorphic setting, the comparison of the Whittaker–Fourier coefficients of the automorphic forms ϑ_{Q_s} and $\vartheta_{\mathrm{spn}(Q_s)}$ show that their difference is not spanned by any form in $r(\chi)$. In modular forms terms, this means this difference lies in the orthogonal complement.

The third part of the Lemma follows from analogous comparison of Whittaker–Fourier coefficients (in the modular form case due to [42]) of the automorphic forms associated to $\theta_{\mathrm{gen}(Q_s)}$ and $\theta_{\mathrm{spn}(Q_s)}$ and by Section 3.1.6. □

3.3. **Equidistribution.** We now state our main equidistribution result, whose proof is given in Section 3.3.3. From now on, let us denote by $w(s) := \#R_s^\times$, for s a supersingular or a superspecial point. Similarly, let us denote by $w(c)$ the cardinality of the endormorphism ring of the reduction of a CM point landing on the component c of the smooth locus $\mathscr{S}^{\mathrm{sm}}_{U,k}$.

Theorem 3.12. *Let $\star \in \{ss, ssp\}$. The sequence of measure $(\mu^\star_{D,c})_{D,c}$ converges, in the weak-* topology, to the canonical measure*

$$\mu^\star(s) := \frac{w(s)^{-1}}{\sum_{s\in\mathscr{S}^\diamond_k} w(s)^{-1}}$$

as D and c vary, that is, as their absolute norms tend to infinity.

Moreover, we also have that the sequence $(\mu^{ssp}_{in,D,c})_{D,c}$ converges to the measure

$$\mu^{ssp}_{in}(c_i) := \frac{w(c_i)^{-1}}{\sum_{j=1}^h w(c_j)^{-1}}$$

on the set $\{c_1, \ldots, c_h\}$ of components of the smooth locus $\mathscr{S}^{sm}_{U,k}$.

3.3.1. *Bounds for the Fourier–Whittaker Coefficients.* The following estimate makes essential use of the Brauer–Siegel theorem, i.e., a lower bound for the class number in the (totally real) number field setting. Note that the ineffectivity of our equidistribution results comes from the use of such notoriously ineffective bound.

Lemma 3.13. *Let $\Gamma_{D,c}$ be the Galois orbit of the CM point defined over the CM field K of discriminant Dc^2. We have that*

$$\frac{r^*(Q_s, Dc^2)\cdot \#\mathscr{O}^\times_{D,c}}{\#\Gamma_{D,c}} = \frac{r^*(\mathrm{spn}(Q_s), Dc^2)\cdot \#\mathscr{O}^\times_{D,c}}{R_K\cdot \#\Gamma_{D,c}} + O(N(Dc^2)^{-\frac{7}{32}+\epsilon}).$$

Proof. Consider an irreducible cuspidal automorphic representation $\overline{\pi}$ of $\widetilde{\mathrm{GL}}_2(\mathbb{A}_F)$ orthogonal to the subspace spanned by theta-representations and pick an automorphic form $\overline{\varphi}$ of $\overline{\pi}$. By [6, Cor.1] the following bound holds

$$a(\overline{\varphi}, m) \ll_{\overline{\varphi},F,\epsilon} N(m)^{\frac{1}{2}-\frac{1}{8}(1-2\theta)}, \tag{3.11}$$

where $a(\overline{\varphi}, m)$ is the m-th normalized Whittaker–Fourier coefficient of $\overline{\varphi}$ and $\theta \in [0, \frac{1}{2}]$ is an approximation towards the Ramanujan–Petersson conjecture. By the work of Kim–Shahidi [33], we can take $\theta = 1/9$. Thus we obtain the exponent $\frac{1}{2} - \frac{1}{8}(1-\frac{2}{9}) = \frac{29}{72}$.

Now, by Proposition 3.11, we have that $\theta_{\mathrm{spn}(Q_s)} - \theta_{Q_s}$ is a cusp form of weight 3/2 which lies in the orthogonal complement of the space of unary theta series. Consider the automorphic representation of half-integral weight spanned by this form, and denote it by $\overline{\vartheta}$. Note that, since $\overline{\vartheta}$ it is not in the image of the correspondence of Lemma 3.3, its Shimura lift $\mathrm{Shi}(\overline{\vartheta})$ is cuspidal. Thus by (3.11) it follows that

$$r^*(\mathrm{spn}(Q_s), Dc^2) - r^*(Q_s, Dc^2) \ll_{\overline{\vartheta},F,\epsilon} |N(Dc^2)|^{\frac{29}{72}+\epsilon}.$$

Furthermore by the Brauer–Siegel theorem (see, for instance, [35, Chap.XVI]) we have the following lower bound for the class number h_K

$$h_K R_K \gg_\epsilon N(Dc^2)^{\frac{1}{2}-\epsilon}$$

where R_K is the regulator of K. Since K is a CM field, by [17, Thm.B] we have that $R_K > 1/4$. Therefore

$$\frac{r^*(Q_s, Dc^2)\cdot \#\mathscr{O}_{D,c}^\times}{\#\Gamma_{D,c}} - \frac{r^*(\mathrm{spn}(Q_s), Dc^2)\cdot \#\mathscr{O}_{D,c}^\times}{\#\Gamma_{D,c}\cdot R_K} \ll N(Dc^2)^{-\frac{7}{32}+\epsilon}.$$

□

For our equidistribution purposes, it is important the trivial remark that the exponent $\frac{29}{72}$ is smaller that the Duke's one $\frac{13}{28}$ as considered in [27, p.518].

3.3.2. *Auxiliary Results.* Let $\left(\frac{K/F}{\cdot}\right)$ denote the Artin symbol.

Lemma 3.14. *Let $h(\mathscr{O}_c)$ and $h(\mathscr{O}_K)$ denote the class numbers of $\mathscr{O}_c$ and $\mathscr{O}_K$ respectively. Then we have*

$$h(\mathscr{O}_c) = h(\mathscr{O}_K)\frac{N(c)}{[\mathscr{O}_K^\times : \mathscr{O}_c^\times]}\prod_{\mathfrak{p}|c}\left(1-\left(\frac{K/F}{\mathfrak{p}}\right)\frac{1}{N(\mathfrak{p})}\right).$$

Proof. We follow the lines of [10, Thm.7.24]. Let $I_K(c)$ denote the group of fractional $\mathscr{O}_K$-ideals prime to c and let $P_{K,\mathscr{O}_F}(c)$ be the subgroup of $I_K(c)$ generated by the principal ideals $\alpha\mathscr{O}_K$, for $\alpha \in \mathscr{O}_K$ such that $\alpha \equiv a \bmod c\mathscr{O}_K$ for $a \in \mathscr{O}_F$ coprime to c. Then, by an immediate adaptation of [10, Prop.7.22], we have

$$h(\mathscr{O}_c) = \#\frac{I_K(c)}{P_{K,\mathscr{O}_F}(c)}.$$

The inclusion $P_{K,\mathscr{O}_F}(c) \subset I_K(c)\cap P_K$ allows us to consider the short exact sequence

$$0 \longrightarrow I_K(c)\cap P_K/P_{K,\mathscr{O}_F}(c) \longrightarrow \mathrm{Cl}(\mathscr{O}_c) \longrightarrow \mathrm{Cl}(\mathscr{O}_K) \longrightarrow 0,$$

so that we just need to compute

$$\#(I_K(f)\cap P_K/P_{K,\mathscr{O}_F}(f)) = h(\mathscr{O}_c)/h(\mathscr{O}_K). \tag{3.12}$$

For $[\alpha] \in (\mathscr{O}_K/c\mathscr{O}_K)^\times$ let us consider the surjective morphism

$$\phi\colon (\mathscr{O}_K/c\mathscr{O}_K)^\times \longrightarrow I_K(c)\cap P_K/P_{K,\mathscr{O}_F}(c)$$

defined by $[\alpha] \mapsto [\alpha\mathscr{O}_K]$. For the sake of simplicity, let us now assume tnat $\#\mathscr{O}_K^\times = 2$. We thus obtain the following short exact sequence

$$1 \longrightarrow (\mathscr{O}_F/c)^\times \xrightarrow{i} (\mathscr{O}_K/c\mathscr{O}_K)^\times \xrightarrow{\phi} I_K(c)\cap P_K/P_{K,\mathscr{O}_F}(c) \longrightarrow 1$$

where i is the natural injection. Analogously to the classical case, the number of units in the quotient $\mathscr{O}_F/c$ is given by

$$\#(\mathscr{O}_F/c)^\times = N(c)\prod_{\mathfrak{p}|c}\left(1-\frac{1}{N(\mathfrak{p})}\right) \tag{3.13}$$

(see [37, Thm.1.19]). Combining formula (3.13) with the Chinese remainder theorem we thus obtain

$$\#(\mathscr{O}_K/c\mathscr{O}_K)^\times = N(c)^2\prod_{\mathfrak{p}|c}\left(1-\frac{1}{N(\mathfrak{p})}\right)\left(1-\left(\frac{K/F}{\mathfrak{p}}\right)\frac{1}{N(\mathfrak{p})}\right) \tag{3.14}$$

where indeed $N(c)^2 = N(c\mathscr{O}_K)$. Note that the Artin symbol summarizes the sign coming from the primes of c being split or inert in K.

Finally, combining (3.14) with (3.13) and (3.14) we obtain

$$h(\mathscr{O}_c)/h(\mathscr{O}_K) = N(c)\prod_{\mathfrak{p}|c}\left(1-\left(\frac{K/F}{\mathfrak{p}}\right)\frac{1}{N(\mathfrak{p})}\right).$$

Lastly, if $\#\mathscr{O}_c^\times > 2$, it is enough to consider the following exact sequence

$$1 \longrightarrow \{\pm 1\} \xrightarrow{j} (\mathscr{O}_F/c)^\times \times \mathscr{O}_K^\times \xrightarrow{\psi} (\mathscr{O}_K/c\mathscr{O}_K)^\times \xrightarrow{\phi} I_K(c)\cap P_K/P_{K,\mathscr{O}_F}(c) \longrightarrow 1$$

where $j\colon \pm 1 \mapsto ([\pm 1], \pm 1)$ and $\psi\colon ([n], u) \mapsto [nu^{-1}]$.
In this case, we have

$$\#(I_K(c)\cap P_K/P_{K,\mathscr{O}_F}(c)) = 2\frac{\#(\mathscr{O}_K^\times/c\mathscr{O}_K)^\times}{\#(\mathscr{O}_F/c)^\times \#\mathscr{O}_K^\times}.$$

Again, by (3.12), we conclude

$$h(\mathscr{O}_c)/h(\mathscr{O}_K) = \frac{N(c)}{[\mathscr{O}_K^\times : \mathscr{O}_c]}\prod_{\mathfrak{p}|c}\left(1-\left(\frac{K/F}{\mathfrak{p}}\right)\frac{1}{N(\mathfrak{p})}\right).$$

□

From now on, let us set $u_{D,c} := \#\mathscr{O}_{D,c}^\times$. We write $u_{D,1}$ to denote the cardinality of the order of trivial conductor.

Corollary 3.15. *We have that the limit defined as*

$$\lim_{k\to\infty}\frac{r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}}$$

exists and it is independent of s and $\mathfrak{p}$.

Proof. The independence of s follows from the definition of $\mathrm{gen}(Q_s)$.

On the other hand, by [28, Thms.49,72,85,86] we have

$$r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k}) = A\frac{h(D\mathfrak{p}^{2k})}{u_{D,\mathfrak{p}^k}} \tag{3.15}$$

where the constant A depends only on the Artin symbol. Combining equation (3.15) with Lemma 3.14 we have

$$r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k}) \sim N(\mathfrak{p}^k)\left(1-\frac{1}{N(\mathfrak{p})}\left(\frac{K/F}{\mathfrak{p}}\right)\right)\frac{h(\mathscr{O}_K)}{u_{D,1}} \tag{3.16}$$

$$\#\Gamma_{D,\mathfrak{p}^k} = N(\mathfrak{p}^k)\left(1-\frac{1}{N(\mathfrak{p})}\left(\frac{K/F}{\mathfrak{p}}\right)\right)\frac{u_{D,\mathfrak{p}^k}}{u_{D,1}}\#\Gamma_{D,1} \tag{3.17}$$

where $\sim$ means equality up to a constant depending on D and the Artin symbol only.

Therefore we obtain

$$\frac{r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}} = C\frac{h(D)}{\#\Gamma_{D,1}}$$

where the right-hand side is independent of $\mathfrak{p}$. □

Lemma 3.16. *Let c be coprime to $\mathfrak{n}$ and ℓ. Then, for every s supersingular or superspecial, in the special fiber at an inert or ramified prime respectively, we have that*

$$f_{D,\mathfrak{p}}(s) := \lim_{k\to\infty}\frac{r^*(Q_s, D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}} = \#R_s^\times \cdot \mu_\diamond, \tag{3.18}$$

where $\diamond \in \{ss, ssp\}$.

Proof. By Section 2.3.6, we have maps to pass from Gross points to supersingular or superspecial points. In view of those constructions, Lemma 3.16 in the supersingular case follows almost verbatim from [44, Thm.1.5] and from Lemma 3.10. On the other hand, the superspecial setting hardly changes the proof. In fact, note that the auxiliary prime ℓ, which one suppose unramified in B, is different from the prime of reduction v, which thus can be of ramification in B. In both cases, the equidistribution is reduced to a classical statement on finite graphs [44, Prop.3.14]. □

Remark 3.17. We invite the reader to note how Vatsal's proof of the previous Lemma is crucial for our results: the absence of any ergodic theory is the reason why we can allow the discriminant to vary in Theorem 3.12, in contrast[20] to [9] and [41].

Next result shows that there is only a single spinor genus.

Lemma 3.18. *Let s be either in the supersingular or in the superspecial locus of $\mathscr{S}_U$. Suppose also that c is coprime to $\mathfrak{n}$ and ℓ. Then*

$$r^*(\mathrm{gen}(Q_s), Dc^2) = r^*(\mathrm{spn}(Q_s), Dc^2).$$

Proof. We proceed as in [27, p.520], i.e., by reductio ad absurdum. Let E be the extension of F of discriminant DD_{Q_s}. Consider a prime $\mathfrak{p}$ coprime to N and ℓ such that $\left(\frac{E/F}{\mathfrak{p}}\right) = -1$. Denote by $W(m)$ the m-th Whittaker–Fourier coefficient of $\theta_{\mathrm{gen}(Q_s)} - \theta_{\mathrm{spn}(Q_s)}$.

By (3.16) we have

$$\begin{aligned} f_{D,\mathfrak{p}}(s) &= \lim_{k\to\infty} \frac{(r^*(\mathrm{spn}(Q_s), D\mathfrak{p}^{2k}) - r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k}))u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}} + \frac{r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}} \\ &= \lim_{k\to\infty} \frac{W(D\mathfrak{p}^{2k})}{\#\Gamma_{D,\mathfrak{p}^k} u^{-1}_{D,\mathfrak{p}^k}} + \frac{r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}}. \end{aligned}$$

By Proposition 3.11 $W(D\mathfrak{p}^{2k})$ is of weight 3/2 and spanned by 1-dimensional theta series as implied by Lemma 3.3. By [43] and following [48, Cor.3.5.3], we can write, in analogy with Remark 3.4, such a Hilbert theta series as

$$\theta_\omega(z) = \sum_m \omega(m)N(m)q^{N(m^2)}$$

where ω is a Hecke character and m an integral ideal. Therefore one obtains the formula

$$W(D\mathfrak{p}^{2k}) = N(\mathfrak{p})^k \left(\frac{E/F}{\mathfrak{p}}\right)^k W(D) \tag{3.19}$$

by the analogous steps of [27, Lem.4.9]. Hence by formula (3.19) and formulae (3.16) we can write

$$N(\mathfrak{p})^k \left(\frac{E/F}{\mathfrak{p}}\right)^k W(D) \cdot \frac{u_{D,1}}{N(\mathfrak{p})^k \left(1 - \frac{1}{N(\mathfrak{p})}\left(\frac{K/F}{\mathfrak{p}}\right)\right) \#\Gamma_{D,1}}.$$

By Lemma 3.16, we have that the limit

$$\lim_{k\to\infty} \frac{r^*(\mathrm{gen}(Q_s), D\mathfrak{p}^{2k})u_{D,\mathfrak{p}^k}}{\#\Gamma_{D,\mathfrak{p}^k}}$$

exists. However, the limit[21]

$$\lim_{k\to\infty} \frac{W(D\mathfrak{p}^{2k})}{\#\Gamma_{D,\mathfrak{p}^k} u^{-1}_{D,\mathfrak{p}^k}} \sim \lim_{k\to\infty} (-1)^k$$

does not exist. This yields the desired contradiction. □

[20] There is a price to pay: Vatsal's equidistribution holds for the reduction at a single prime only.

[21] Where the symbol $\sim$ means equality up to a constant.

3.3.3. *Proof of Theorem 3.12.* Combining the subconvexity bound of Lemma 3.13 with Lemma 3.18, we obtain

$$\frac{r^*(Q_s, Dc^2)u_{D,c^2}}{\#\Gamma_{D,c}} = \frac{r^*(\mathrm{gen}(Q_s, Dc^2))u_{D,c}}{R_K \cdot \#\Gamma_{D,c}} + O(N(Dc^2)^{-\frac{7}{32}+\epsilon}).$$

Let s be a geometric point either in the supersingular or in the superspecial locus. By Lemma 3.10 we have that the number of optimal embeddings of $\mathscr{O}_{D,c}$ into R_s is given by

$$r^*(Q_s, Dc^2)\frac{u_{D,c}}{w_s}. \tag{3.20}$$

Since the sum over all geometric points s of (3.20) is equal to $\#\Gamma_{D,c}$, we obtain

$$\begin{aligned} 1 &= \sum_{s\in\mathscr{S}_k^\diamond} \frac{r^*(Q_s, Dc^2)u_{D,c}}{w_s \cdot \#\Gamma_{D,c}} \\ &= \frac{r^*(\mathrm{gen}(Q_s, Dc^2))u_{D,c}}{R_K \cdot \#\Gamma_{D,c}} \sum_{s\in\mathscr{S}_k^\diamond} \frac{1}{w_s} + O(N(Dc^2)^{-\frac{7}{32}+\epsilon}), \end{aligned}$$

where the independence of $r^*(\mathrm{gen}(Q_s, Dc^2))$ from s gives the second equality.

Since

$$\frac{r^*(\mathrm{gen}(Q_s, Dc^2))u_{D,c}}{R_K \cdot \#\Gamma_{D,c}} = \frac{1}{\sum_{s\in\mathscr{S}_k^\diamond} w_s^{-1}} + O(N(Dc^2)^{-\frac{7}{32}+\epsilon})$$

we have that $\lim_{N(Dc^2)\to\infty} \dfrac{r^*(\mathrm{gen}(Q_s, Dc^2))u_{D,c}}{R_K \cdot \#\Gamma_{D,c}}$ exists.

Therefore

$$\begin{aligned} &\lim_{N(Dc^2)\to\infty} \frac{r^*(Q_s, Dc^2)u_{D,c}}{\#\Gamma_{D,c}} \\ &= \lim_{N(Dc^2)\to\infty} \frac{r^*(\mathrm{gen}(Q_s, Dc^2))u_{D,c}}{R_K \cdot \#\Gamma_{D,c}} + O(N(Dc^2)^{-\frac{7}{32}+\epsilon}) \\ &= \frac{1}{\sum_{s\in\mathscr{S}_k^\diamond} w_s^{-1}} = w_s\mu^\star \end{aligned}$$

and so we conclude.

4. An André–Oort-like Result

Following [39] and [40], we consider the 2-dimensional arithmetic André–Oort conjecture [40, Conj.2.3] for the integral model $\mathscr{S}$ over $\mathbb{Z}$. For the general statement of this conjecture for Shimura varieties, see [40, Sec.2.1].

4.0.1. *Horizontal André–Oort in Pencils.* Intuitively, the word "horizontal" refers to the fact that we allow the characteristic p to vary over $\mathbb{Z}$.

Let $\mathcal{S}(d_K)^{\mathrm{CM}}$ denote the set of CM points with discriminant d_K. Note that two CM points have the same discriminant if and only if they belong to the same $G_{\mathbb{Q}}$-orbit. A *special* Cartier divisor S is an element of the group of divisors $\mathrm{Div}(\mathscr{S}_{\overline{\mathbb{F}}_p})$ of the form

$$\sum_{z\in\mathcal{S}(d_K)^{\mathrm{CM}}} [red_p(z)].$$

Note that the degree of S coincides with its cardinality $\#S$.

Consider the Galois orbit of CM points of discriminant d_K. Denote by E be a field and let $s = \operatorname{Spec}\overline{E}$. For a E-scheme X locally of finite type, we recall that there is an equivalence between scheme theoretic points of X and Galois orbits of geometric points. Namely, the map

$$X(\overline{E}) \longrightarrow X,\ x \mapsto x(s)$$

induces a bijection between the set of G_E-orbits in $X(\overline{E})$ and the set of closed points in X. Therefore such a $G_{\mathbb{Q}}$-orbit corresponds to a unique closed point $x(d_K)$ in $\mathcal{S}$ over $\mathbb{Q}$. Note that the Zariski closure of $x(d_K)$, which we call *special* curve, is such that $\overline{x(d_K)}^{\mathrm{Zar}}(\overline{\mathbb{F}}_p) = red_p(\mathcal{S}(d_K)^{\mathrm{CM}})$.

This horizontal extension of the André–Oort conjecture (as proposed in [39, Conj.1]) concerns the Zariski closure of a collection of special divisors in $\mathscr{S}$.

Theorem 4.1. *Let $\mathbb{S}$ denote a collection of special divisors. Then the Zariski closure of $\cup\{S : S \in \mathbb{S}\}$ in $\mathscr{S}$ is a finite union of the following subsets:*

(1) *a special divisor S, for p and d_K fixed;*
(2) *the special curve $\overline{x(d_K)}^{Zar}$, for d_K fixed;*
(3) *the fibers $\mathscr{S}_{\mathbb{F}_p}$ over p, for p fixed;*
(4) *the integral model $\mathscr{S}$.*

As in [40, Def.3.1], we call the *characteristic* of a special divisor S the prime p such that $S \subset \mathscr{S}_{\overline{\mathbb{F}}_p}$. The structure of a special divisor S falls under one of the following cases:

(1) if B_p is ramified, it consists
   - either of superspecial points;
   - or of supersingular points;
(2) if B_p is unramified, it consists
   - either of supersingular points;
   - or of ordinary points.

We thus label a special divisor as *superspecial*, *supersingular* or *ordinary* if it lies either in the superspecial or in supersingular, or in the ordinary locus of the special fiber of $\mathscr{S}$. Note that such a divisor will be entirely contained in only one of these sets.

We conclude with a result which adapts [32, Thm.12.4.5] to our Shimura curves setting and that we are gonna exploit in the next final session.

As in [29, p.364], we recall that a $\mathscr{O}_B \otimes \mathbb{Z}_p$-module N decomposes as $N = N_1 \oplus N_2 = N_1 \oplus (N_{2,1} \oplus N_{2,2})$, where the $M_2(\mathbb{Z}_p)$-module N_2 decomposes[22] into the sum of two $\mathbb{Z}_p$-modules.
Consider the following moduli problem $\mathcal{F}$ of $\mathbb{Z}_p$-algebras

$$S \longmapsto [A, \theta, \kappa]$$

where the triple $[A, \theta, \kappa]$ is an isomorphism class consisting of

(1) an abelian surface A over S with an action $\iota\colon \mathscr{O}_B \hookrightarrow \mathrm{End}_S(A)$ such that $\mathrm{Lie}(A)_{2,1}$ is of rank-1 and $\mathbb{Z}_p$ acts on it[23];
(2) a class of polarizations $\theta\colon A \to A^\vee$ such that, for every $b \in \mathscr{O}_B$, the associated Rosati involution takes $\iota(b)$ to $\iota(b^*)$;
(3) a class of $\mathscr{O}_B$-linear rigidifications $\kappa\colon \mathscr{O}_B \otimes \widehat{\mathbb{Z}} \simeq \prod_p T_p(A)$.

This moduli problem is representable by $\mathscr{S} \times \operatorname{Spec} \mathbb{Z}_p$. For more details, we refer to [29, Sec.4.1] and [48, Prop.1.1.5].

Let us consider the category of abelian surfaces base-changed to $\overline{\mathbb{F}}_p$ and the moduli problem $\overline{\mathcal{F}}$ on it defined by base-change over $\overline{\mathbb{F}}_p$. Indeed, $\overline{\mathcal{F}}$ is representable by the special fiber $\mathscr{S} \times \operatorname{Spec} \overline{\mathbb{F}}_p$.

Lemma 4.2. *The number of supersingular points in the special fiber of $\mathscr{S}$ at p is given by*

$$\#\mathscr{S}^{ss}_{\overline{\mathbb{F}}_p} = \frac{p-1}{24} \deg(\overline{\mathcal{F}}),$$

where $\deg(\overline{\mathcal{F}})$ is the degree with which $\overline{\mathcal{F}}$ is finite étale over the stack of abelian surfaces over $\overline{\mathbb{F}}_p$.

[22] After choosing an idempotent of $M_2(\mathbb{Z}_p)$.
[23] Since $\mathrm{Lie}(A)$ is a $\mathscr{O}_B \otimes \mathbb{Z}_p$-module, it admits the above decomposition.

Proof. Let us refresh a few facts on the Hasse invariant H for Shimura curves, following [29, Sec.6]. Firstly, consider the universal abelian scheme $\varepsilon\colon \mathscr{A}\to\mathscr{S}$ as in Section 2.3.1, and let $\Omega^1_{\mathscr{A}/\mathscr{S}}$ be the canonical bundle. Note that the $\mathscr{O}_{\mathscr{S}}$-module $\varepsilon_*\Omega^1_{\mathscr{A}/\mathscr{S}}$ is also a $\mathscr{O}_{B_p}$-module, and we introduce $\underline{\omega}=(\varepsilon_*\Omega^1_{\mathscr{A}/\mathscr{S}})_{2,1}$, which is a line bundle[24] over $\mathscr{S}$.

We now have that H is a modular form of weight $p-1$ over $\mathscr{S}\otimes\overline{\mathbb{F}}_p$ constructed as a section of bundle $\underline{\omega}^{\otimes(p-1)}$ over $\mathscr{S}\times\operatorname{Spec}\overline{\mathbb{F}}_p$. By [29, Prop.6.1,6.3] it follows that H vanishes exactly in the supersingular locus and it has only simple zeros. Therefore, by [32, Cor.10.13.12], the number of supersingular points, i.e., the number of zeros of H counted with multiplicities, is equal to

$$\deg\underline{\omega}^{\otimes(p-1)}=(p-1)\deg\underline{\omega}=\frac{p-1}{24}\deg(\overline{\mathcal{F}}).$$

□

4.0.2. *Proof of Theorem 4.1.* Armed with the constructions and results of the rest of this paper, we follow the line of reasoning of [40]. Note that what follows holds, mutatis mutandis, also for a Shimura curve $\mathcal{S}_U$ attached to $B=M_2(\mathbb{Q})$, i.e., for modular curves with arbitrary level structure.

We begin with parts (1) and (3) of Theorem 4.1. Since the fiber of the map $\mathscr{S}\to\operatorname{Spec}\mathbb{Z}$ above a prime p is the curve $\mathscr{S}\times\operatorname{Spec}\mathbb{F}_p$, then one easily concludes, because the special fiber of $\mathscr{S}$ at p is of dimension one. It thus follows that there is either a finite set of special points, so the closure is the just their union, or the closure of the (infinitely many) special points is $\mathscr{S}_{\mathbb{F}_p}$ itself, because the closure of any infinite subset of an irreducible curve is the whole curve.

Concerning the case (2) of Theorem 4.1, we reduce to the case of special divisors of bounded degree and characteristic going to infinity. Moreover, we can also reduce to the case of $\overline{\cup\{S : S\in\mathbb{S}\}}^{\mathrm{Zar}}\to\operatorname{Spec}\mathbb{Z}$ sending its generic points to the generic point of $\operatorname{Spec}\mathbb{Z}$. To show this, let Z' denote the union of the irreducible components of the Zariski closure of $\cup\{S : S\in\mathbb{S}\}$ which are contained in a special fiber $\mathscr{S}_{\mathbb{F}_p}$. Since the characteristic is not bounded, in Z' there are a finite number of special divisor. We thus consider Z to be Z' minus such divisors. Thus $Z\to\operatorname{Spec}\mathbb{Z}$ sends the generic points of $\overline{\cup\{S : S\in\mathbb{S}\}}^{\mathrm{Zar}}$ to the generic point of $\operatorname{Spec}\mathbb{Z}$.

Let p be unramified in B and split in K, i.e., the associated special divisor S is ordinary. We say that S is *canonical* if it is the special fiber at p of a special curve $\overline{x(d_K)}^{\mathrm{Zar}}$ such that $\#red_p(\mathcal{S}(d_K)^{\mathrm{CM}})=\#\mathcal{S}(d_K)^{\mathrm{CM}}$. By Serre–Tate theory (see [7, Sec.0.9]), there is a unique such canonical lifting $x(d_K)$. For a detailed account on this beautiful theory, we invite the reader to go through [31].

As in [40, Sec.3.2.4.1], there are only finitely many special curves of bounded degree. Moreover, the ordinary special divisors lift to special curves with bounded discriminant, so that we have only a finite number of discriminants and we deal with a finite union of special curves. This case is 1-dimensional, so we conclude as for parts (1) and (3).

Let p be unramified in B and inert in K, i.e., the associated S is supersingular.

As in [40, Thm.3.8], this case reduces to showing that

$$\lim_{p\to\infty}\lim_{d_K\to-\infty}\#red_p(\mathcal{S}(d_K)^{\mathrm{CM}})=\infty,\tag{4.1}$$

for p coprime to the conductor. We also need to impose that $p\to\infty$ for supersingular special divisor. In fact, for a fixed p,

$$\lim_{d_K\to-\infty}\#red_p(\mathcal{S}(d_K)^{\mathrm{CM}})=\#\mathscr{S}^{\mathrm{ss}}_{\overline{\mathbb{F}}_p}=\frac{p-1}{24}\deg(\overline{\mathcal{F}})$$

where the first equality is a consequence of Theorem 3.12 the cardinality of the supersingular locus comes from Lemma 4.2.

[24] This follows from the fact that $\mathrm{Lie}(\mathscr{A})_{2,1}$ is locally free of rank-1.

In order to prove (4.1), the rather general approach of [40] applies also to our scenario. Namely, this consists of finding some bounds for $\#red_p(\mathcal{S}(d_K)^{\mathrm{CM}})$. In order to give such bounds, the first step consists of noticing that one can easily reduce to give a uniform bound to the Fourier coefficients of the theta form attached to the ternary quadratic form Q_s, for s a supersingular point as in Section 3.2.2. To do so, the toolbox in [40, Sec.5,6] contains, in order, the Dirichlet–Hermite bound, a more convoluted subconvexity bound, which involves the representation numbers associated to the genus $\mathrm{gen}(Q_s)$ as in Section 3.2.6 and the so-called "slices" method. By moving the problem to bounding the representation numbers, these methods apply verbatim to our setting.

In the case of a superspecial special divisor, the proof follows the very same lines as in the superspecial case. A (minor) difference is the cardinality of the superspecial locus: by [30, Sec.5] we have that $\#\mathscr{S}_{\mathbb{F}_p}^{\mathrm{ssp}}$ grows polynomially in p. We also finally point out that for the Gross lattice, one has to consider the construction we described in Sections 2.3.4 and 3.2.2 for superspecial points.

In conclusion, let us consider the case of $\mathbb{S}$ containing an infinite subsequence with both the characteristic and the degree going to infinity, i.e., the case (4) of Theorem 4.1. If the characteristics of the special points do not belong to a finite set, then their closure Z intersects infinitely many fibers, and the divergence of the degrees implies that

$$\limsup_{p\to\infty} \#(Z \cap \mathscr{S}_{\mathbb{F}_p}) = \infty,$$

namely, Z has intersection of arbitrarily large order as p goes to infinity. Now we claim that a closed (hence proper) subscheme Z of $\mathscr{S}$ have bounded intersection with almost all fibers. Therefore we conclude that $Z = \mathscr{S}$. To prove such a claim, it is enough to note that Z does not contain all of $\mathscr{S}_{\mathbb{Q}} = \mathcal{S}$, hence $Z_{\mathbb{Q}}$ is finite. By generic flatness, there is some $N > 0$ such that Z is flat over $\mathbb{Z}[1/N]$. Therefore we have that $\#(Z \times \operatorname{Spec} \mathbb{F}_p) = \#Z_{\mathbb{Q}}$ over $\mathbb{Z}[1/N]$. This implies that the size of $Z_{\mathbb{F}_p}$ is bounded for all p not dividing N.

References

[1] Aka, Menny; Luethi, Manuel; Michel, Philippe; Wieser, Andreas. *Simultaneous supersingular reductions of CM elliptic curves.* Journal für die reine und angewandte Mathematik (Crelles Journal), vol. 2022, no. 786, 2022, pp. 1-43. 1

[2] Arthur, James. *Eisenstein series and the trace formula.* Automorphic forms, representations and L-functions (Pro. Sympos. Pure Math., Oregon State Univ., Corvallis, ore., 1977), Part 1, pp. 253-274, Proc. Sumpos. Pure Math, XXXIII, Amer. Math. Soc., Providence, R.I., 1979. 17

[3] Boutot, Jean-François; Carayol, Henri. *Uniformisation p-adique des courbes de Shimura: les théorémes de Cerednik et de Drinfeld.* Courbes modulaires et courbes de Shimura (Orsay, 1987-1988). Astérisque No. 196-197 (1991), 7, 45-158 (1992). 9

[4] Boutot J.-F.; Zink Thomas. *The p-adic Uniformization of Shimura Curves.* Arxiv preprint. 3

[5] Bertolini, Massimo; Darmon Henri. *Heegner points on Mumford-Tate curves.* Invent. Math. 126 (1996), no. 3, 413-456. 5

[6] Blomer, Valentin; Harcos, Gergely. *Twisted L-functions over number fields and Hilbert's eleventh problem.* Geom. Funct. Anal. 20 (2010), no. 1, 1-52. 4, 23

[7] Carayol, H. *Sur la mauvaise réduction des courbes de Shimura.*(French) Compositio Math.59(1986), no.2, 151-230. 3, 5, 8, 29

[8] Cornut, Christophe; Jetchev, Dimitar. *Liftings of reduction maps for quaternion algebras.* Bull. Lond. Math. Soc.45(2013), no.2, 370-386. 11, 12

[9] Cornut, C.; Vatsal, V. *CM points and quaternion algebras.* Doc. Math. 10 (2005), 263-309. 1, 2, 10, 26

[10] Cox, D *Primes of the form* $x^2 + y^2 = 1$. Wiley, New York, 1989. 24

[11] Deligne, Pierre. *Travaux de Shimura.* Séminaire Bourbaki, Exp. No. 389, pp. 123-165. Lecture Notes in Math., Vol. 244, Springer, Berlin, 1971.

[12] Disegni, Daniel. *The p-adic Gross–Zagier formula on Shimura curves.* Compositio Mathematica. 2017;153(10):1987-2074. 13

[13] Disegni, D. *p-adic equidistribution of CM points.* Comment. Math. Helv. 97 (2022), no. 4, 635-668. 2

[14] Drinfeld, Vladimir G. *Coverings of p-adic symmetric domains.* (Russian) Funkcional. Anal. i Prilozhen. 10 (1976), no. 2, 29-40. 9

[15] Elkies, Noam; Ono, Ken; Yang, Tonghai. *Reduction of CM elliptic curves and modular function congruences.* Int. Math. Res. Not.(2005), no.44, 2695-2707. 3

[16] Flicker, Yuval Z. *Automorphic forms on covering groups of* GL(2). Invent. Math. 57 (1980), no. 2, 119-182. 17

[17] Friedman, Eduardo. *Analytic formulas for the regulator of a number field.* Invent. Math.98(1989), no.3, 599-622. 24

[18] Garrett, Paul B. *Holomorphic Hilbert modular forms.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990. xvi+304 pp. 16, 18

[19] Garrett, Paul. *Modern analysis of automorphic forms by example.* Vol. 1. Cambridge Stud. Adv. Math., 173 Cambridge University Press, Cambridge, 2018. 16, 18

[20] Gelbart Stephen. *Weil's Representation and the Spectrum of the Metaplectic Group.* Lecture Notes in Math., Vol. 530 Springer-Verlag, Berlin-New York, 1976. 14, 15, 17, 20

[21] Gelbart, S.; Piatetski-Shapiro, Ilya. *Distinguished Representations and Modular Forms of Half-integral Weight* Invent. Math. 59 (1980), no. 2, 145-188. 3, 15, 16, 17

[22] Gelbart, S., Piatetski-Shapiro, I. *On Shimura's correspondence for modular forms of half-integral weight.* Proc. Int. Coll. Auto. Forms, Rep. Theory and Arith., Bombay 1980. 3, 13, 16, 17

[23] Gross, Benedict. *Heights and the special values of L-series.* Number Theory (Montreal, Que., 1985). In: CMS Conference Proceedings, vol. 7, pp. 115-187. American Mathematical Society, Providence, 1987. 3, 5, 6, 7, 9, 19, 21

[24] Hanke, Jonathan. *Local densities and explicit bounds for representability by a quadratic form.* Duke Math. J. 124(2): 351-388, 2004. 16, 23

[25] Hanke, J. *Quadratic Forms and Automorphic Forms.* Dev. Math., 31 Springer, New York, 2013, 109-168. 20, 21

[26] Hiraga Kaoru, Ikeda Tamotsu. *On the Kohnen plus space for Hilbert modular forms of half-integral weight I.* Compositio Mathematica. 2013;149(12):1963-2010. 20

[27] Jetchev, Dimitar; Kane, Ben. *Equidistribution of Heegner points and ternary quadratic forms.* Math. Ann. 350 (2011), no. 3, 501-532. 1, 2, 3, 7, 10, 20, 21, 24, 26

[28] Jones, Burton. *The arithmetic theory of quadratic forms.* Carcus Monograph Series, no. 10. The Mathematical Association of America, (Buffalo) 1950. 21, 25

[29] Kassaei, Payman L.; *$\mathcal{P}$-adic modular forms over Shimura curves over totally real fields.* Compositio Mathematica. 140 (2) (2004), pp. 359-395. 28, 29

[30] Katsura, Toshiyuki; Oort, Frans. *Families of supersingular abelian surfaces.* Compositio Mathematica, 62 (2) (1987), pp. 107-167. 30

[31] Katz, Nicholas M.; *Serre–Tate local moduli.* In: Giraud, J., Illusie, L., Raynaud, M. (eds) Surfaces Algébriques. Lecture Notes in Mathematics, vol 868. Springer, 1981. 29

[32] Katz, N.M.; Mazur, Barry. *Arithmetic moduli of elliptic curves.* Ann. of Math. Stud., 108 Princeton University Press, Princeton, NJ, 1985. 8, 28, 29

[33] Kim, Henry H.; Shahidi, Freydoon. *Cuspidality of symmetric powers with applications.* Duke Math. J.112(2002), no.1, 177-197. 23

[34] Labesse, Jean-Pierre *The Langlands spectral decomposition.* London Math. Soc. Lecture Note Ser., 467 Cambridge University Press, Cambridge, 2021, 176-214. 17, 18

[35] Lang, Serge. *Algebraic number theory.* Second edition, Grad. Texts in Math., 110 Springer-Verlag, New York, 1994. 24

[36] Michel, Philippe. *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points.* Ann. of Math. (2) 160 (2004), no. 1, 185-236. 1

[37] Narkiewicz, W. *Elementary and Analytic Theory of Algebraic Numbers.* (2nd edition), Springer, 1990. 24

[38] Raghunatan A., Tanabe Naomi. *Notes on the arithmetic of Hilbert modular forms.* J. Ramanujan Math. Soc.26(2011), no.3, 261319. 13, 14

[39] Richard, Rodolphe. *Problemes de type André-Oort en pinceau arithmétique.* Exp. Math.,vol. 41,2023, pp. 618-630. 4, 27, 28

[40] Richard, R. *A two-dimensional arithmetic André-Oort problem.* Bull. Lond. Math. Soc.55(2023), no.3, 1459-1488. 4, 27, 28, 29, 30

[41] Saettone, Francesco Maria. *Equidistribution of CM points on a Shimura curve modulo a ramified prime.* Arxiv preprint. 1, 2, 4, 9, 10, 26

[42] Schulze-Pillot, Rainer. *Thetareihen positiv definiter quadratischer Formen.* Invent. Math. 75 (1984), 283-299. 23

[43] Shimura, Goro. *On Hilbert modular forms of half-integral weight.* Duke Math. J. 55(4), 765-838 (1987). 12, 18, 26

[44] Vatsal, Vinayak. *Uniform distribution of Heegner points.* Invent. Math. 148 (2002), no. 1, 1-46. 26

[45] Voight, John. *Quaternion algebras.* Graduate Texts in Mathematics, 288. Springer, 2021. 21

[46] Yuan, Xinyi; Zhang, Shou-Wu; Zhang, Wei. *The Gross-Zagier formula on Shimura curves.* Annals of Mathematics Studies, 184. Princeton University Press, Princeton, NJ, 2013. x+256 pp. 10

[47] Zhang, Shouwu. *Gross-Zagier formula for* GL_2. Asian J. Math. 5 (2001). no. 2, 183-290. 2, 7, 8, 9

[48] Zhang, Shouwu. *Heights of Heegner points on Shimura curves.* Ann. of Math. (2) 153 (2001), no. 1, 27-147. 5, 6, 8, 13, 26, 28

[49] Zhang, Xiaoyu. *Simultaneous supersingular reductions of Hecke orbits.* Arxiv preprint. 1

Department of Mathematics, Ben-Gurion University of the Negev, Israel
Email address: saettone@post.bgu.ac.il