

# CM-POINTS AND LATTICE COUNTING ON ARITHMETIC COMPACT RIEMANN SURFACES

MONTSERRAT ALSINA AND DIMITRIOS CHATZAKOS

ABSTRACT. Let  $\mathcal{X}(D, 1) = \Gamma(D, 1) \backslash \mathbb{H}$  denote the Shimura curve of level  $N = 1$  arising from an indefinite quaternion algebra of fixed discriminant  $D$ . We study the discrete average of the error term in the hyperbolic circle problem over Heegner points of discriminant  $d < 0$  on  $\mathcal{X}(D, 1)$  as  $d \rightarrow -\infty$ . We prove that if  $|d|$  is sufficiently large compared to the radius  $r \approx \log X$  of the circle, we can improve on the classical  $O(X^{2/3})$ -bound of Selberg. Our result extends the result of Petridis and Risager for the modular surface to arithmetic compact Riemann surfaces.

## 1. INTRODUCTION

Let  $\mathbb{H}$  denote the hyperbolic plane and  $\rho(z, w)$  the hyperbolic distance of two fixed points  $z, w \in \mathbb{H}$ . For  $\Gamma$  a cocompact or a general cofinite Fuchsian group, the hyperbolic lattice counting problem asks to estimate the quantity

$$(1.1) \quad N(X; z, w) = \# \left\{ \gamma \in \Gamma : \rho(z, \gamma w) \leq \cosh^{-1} \left( \frac{X}{2} \right) \right\},$$

as  $X \rightarrow \infty$ . The study of this problem was initiated in [10]. Further, the asymptotic behaviour of  $N(X; z, w)$  was studied in [12, 13, 19, 20, 32, 37] using tools from the spectral theory of automorphic forms in  $L^2(\Gamma \backslash \mathbb{H})$ . Let  $\Delta$  be the Laplacian of the hyperbolic surface  $\Gamma \backslash \mathbb{H}$  and  $\{u_j\}_{j=0}^{\infty}$  be an  $L^2$ -normalized sequence of eigenfunctions of  $-\Delta$  (Maass forms) with eigenvalues  $\{\lambda_j\}_{j=0}^{\infty}$ . We write  $\lambda_j = s_j(1 - s_j) = 1/4 + t_j^2$ . The asymptotics of  $N(X; z, w)$  can be deduced as an application of the spectral theorem for  $L^2(\Gamma \backslash \mathbb{H})$  for appropriately chosen automorphic kernels (also called the pre-trace formula). Selberg [37] first proved that, as  $X \rightarrow \infty$ ,

$$(1.2) \quad N(X; z, w) \sim M(X; z, w) := \sum_{1/2 < s_j \leq 1} \sqrt{\pi} \frac{\Gamma(s_j - 1/2)}{\Gamma(s_j + 1)} u_j(z) \overline{u_j(w)} X^{s_j}$$

and

$$(1.3) \quad E(X; z, w) := N(X; z, w) - M(X; z, w) = O(X^{2/3}).$$

For the error term  $E(X; z, w)$  we have the conjecture

$$(1.4) \quad E(X; z, w) = O_{\epsilon}(X^{1/2+\epsilon})$$

for every  $\epsilon > 0$  [32, 35]. We refer to [7, 9, 17] for upper bounds on the second moments of the error term and to [8, 35] for mean-value and  $\Omega$ -results supporting this conjecture.

More information can be obtained for the error term when the group  $\Gamma$  is arithmetic. For  $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ , Petridis and Risager recently studied the average growth of  $E(X; z, z)$  on compact subsets of the modular surface  $\Gamma \backslash \mathbb{H}$  both in the local [33] and the discrete aspect [34]. To state their discrete average result, for  $d < 0$  a fundamental discriminant denote by  $\Lambda_d$  the set of CM (or Heegner) points

---

*Date:* January 16, 2020.

*2010 Mathematics Subject Classification.* Primary 11F72; Secondary 37C35, 37D40.

of discriminant  $d$  and by  $h(d)$  the class number  $h(d) = \#\Lambda_d$ . If  $f \geq 0$  is a smooth compactly supported function on  $\Gamma \backslash \mathbb{H}$ , Duke's equidistribution theorem states that

$$(1.5) \quad \frac{1}{h(d)} \sum_{z \in \Lambda_d} f(z) \rightarrow \frac{1}{\text{vol}(\Gamma \backslash \mathbb{H})} \int_{\Gamma \backslash \mathbb{H}} f(z) d\mu(z),$$

as  $d \rightarrow -\infty$ . Applying Duke's theorem and their local average result for the error term over compact subsets of the modular surface [33] they proved the following upper bound, which improves on average on Selberg's bound for  $|d| \geq X^{11/2+\epsilon}$ .

**Theorem 1.1** (Petridis-Risager [34]). *Let  $f$  be a non-negative smooth compactly supported function on the modular surface. Then, as  $X \rightarrow \infty$ ,*

$$(1.6) \quad \frac{1}{h(d)} \sum_{z \in \Lambda_d} f(z) E(X; z, z) = O_{f,\epsilon}(X^{7/12+\epsilon} + X^{4/5+\epsilon} |d|^{-4/165+\epsilon}).$$

In this paper we study the discrete average of the error term  $E(X; z, z)$  on the indefinite Shimura curves  $\mathcal{X}(D, N) = \Gamma(D, N) \backslash \mathbb{H}$ . Here  $\Gamma(D, N)$  is the cocompact discrete subgroup of  $\text{SL}_2(\mathbb{R})$  of level  $N$  arising from an indefinite quaternion algebra

$$H = \left( \frac{a, b}{\mathbb{Q}} \right)$$

of discriminant  $D := D_H$  (see subsection 2.1). Let also  $\Lambda_{D, N, d}$  denote the set of CM points on  $\mathcal{X}(D, N)$  of fundamental discriminant  $d < 0$  (when non-empty) and  $h(D, N, d) = \#\Lambda_{D, N, d}$  the class number counting the number of CM points on  $\mathcal{X}(D, N)$ . For simplicity reasons, for the rest of the paper we restrict to level  $N = 1$  and we fix the notation  $\Lambda_{D, d} = \Lambda_{D, 1, d}$  and  $h(D, d) := h(D, 1, d)$ . Applying Duke's equidistribution theorem for arithmetic compact surfaces, we prove the following theorem.

**Theorem 1.2.** *Let  $H$  be an indefinite quaternion algebra over  $\mathbb{Q}$  of discriminant  $D$  and  $\Gamma(D, 1)$  the arithmetic cocompact Fuchsian group of level 1 arising from  $H$ . Let  $f$  be a non-negative smooth function on the Shimura curve  $\mathcal{X}(D, 1) = \Gamma(D, 1) \backslash \mathbb{H}$ . Then*

$$(1.7) \quad \frac{1}{h(D, d)} \sum_{z \in \Lambda_{D, d}} f(z) E(X; z, z) = O_{f, D, \epsilon}(X^{5/8+\epsilon} + X^{4/5+\epsilon} |d|^{-2/85+\epsilon})$$

as  $X \rightarrow \infty$ .

Here, the limit is considered over  $d$ 's such that the class number is nonzero. As a corollary, our result improves on Selberg's bound for arithmetic compact surfaces when the discriminant satisfies  $|d| \geq X^{\frac{34}{5}+\epsilon}$ . For  $|d| \geq X^{\frac{119}{16}+\epsilon}$  we get an upper bound  $O(X^{5/8+\epsilon})$ .

A main ingredient in the proof of Theorem 1.2 is the recent local average result of Biro [4] for general Riemann surfaces of finite area. This is the first reason our upper bound is slightly weaker from that of Theorem 1.1. A second reason is we have not tried to optimize the exponents in all the possible aspects, for instance the bound (2.9) on average, as our main goal is to indicate how the method of [33] can be combined with the spectral result of Biro [4] to deduce a bound for the discrete average of  $E(X; z, z)$  (see section 2 for more details). In section 2 we discuss the necessary background and the tools in the proof of Theorem 1.2 (equidistribution and bounds for Weyl sums, Biro's local average result, bounds for inner products of eigenfunctions). We give the proof of our result in section 3.

**1.1. Acknowledgments.** We are grateful to the School of Mathematics at the University of Bristol, where this work initially started, for the hospitality and the friendly environment. The first author is supported by Grant MTM2015-63829-P and Grant MTM2015-66716-P. The second author would also like to thank the Mathematics department of King's College London, for the support during the academic year 2016-17. This work was partially supported by the LMS 150th Anniversary Postdoctoral Mobility Grant 2016-17 and the European Unions Seventh Framework Programme (FP7/2007-2013)

/ ERC grant agreements no 335141 Nodal. Currently, he is supported by a Labex CEMPI (ANR-11-LABX-0007-01). We would like to thank P. Bayer for useful comments, V. Blomer, Ph. Michel, P. Nelson, R.M. Nunes for useful comments on subconvexity and Y. Petridis for his valuable advice.

## 2. BACKGROUND AND AUXILLARY RESULTS

In this section we summarize the necessary background about the arithmetic compact Riemann surfaces  $\mathcal{X}(D, N)$  and the spectral theory of automorphic forms on them.

**2.1. Background on compact arithmetic groups.** We refer to [23, p. 301], [24, p. 385] as a basic reference for quaternionic groups and to [1] for a detailed study of their arithmetic theory. For  $a, b$  square-free integers with  $a > 0$  let

$$(2.1) \quad H = \left( \frac{a, b}{\mathbb{Q}} \right)$$

be the indefinite quaternion algebra over  $\mathbb{Q}$  linearly generated by  $\{1, \omega, \Omega, \omega\Omega\}$ , where  $\omega^2 = a$ ,  $\Omega^2 = b$  and  $\omega\Omega + \Omega\omega = 0$ . We say that  $H$  ramifies at  $p$  if  $H_p = H \otimes_{\mathbb{Q}} \mathbb{Q}_p$  is a division algebra. Let  $N(x)$  denote the norm of  $x \in H$ . For  $R$  a maximal order of  $H$ , denote by  $R(n)$  the elements  $x \in R$  of norm  $n$ . Denote also by  $\phi$  the embedding that maps

$$(2.2) \quad x = x_0 + x_1\omega + x_2\Omega + x_3\omega\Omega = \xi + \eta\Omega$$

to

$$(2.3) \quad \phi(x) = \begin{pmatrix} \bar{\xi} & \eta \\ b\bar{\eta} & \xi \end{pmatrix}.$$

Then, the group

$$(2.4) \quad \Gamma_H := \phi(R(1)) \subset \mathrm{SL}_2(\mathbb{R})$$

is a Fuchsian group of the first kind, and the quotient  $\Gamma_H \backslash \mathbb{H}$  is a Riemann surface. The discriminant  $D = D_H$  of the algebra is defined as the product of the ramified primes of  $H$ .

Now let  $D, N$  be natural numbers such that: i)  $D$  is a product of an even number of different primes and, ii)  $N > 1$  is a natural number with  $(D, N) = 1$ . If  $R = \mathcal{O}(D, N)$  is an Eichler order of level  $N$  and  $\phi$  a monomorphism  $: H \rightarrow M_2(\mathbb{R})$ , we denote the group  $\Gamma_H$  associated to  $\mathcal{O}(D, N)$  and  $\phi$  by  $\Gamma(D, N)$ . We have

$$(2.5) \quad \Gamma(D, N) \subset \mathrm{SL}_2(\mathbb{Q}(\sqrt{a})).$$

The theory of Shimura provides a canonical model  $\mathcal{X}(D, N)$  for  $\Gamma(D, N) \backslash \mathbb{H}$  and a modular interpretation. The canonical model  $\mathcal{X}(D, N)$  is a projective curve defined over  $\mathbb{Q}$ . The quotient  $\Gamma(D, N) \backslash \mathbb{H}$  is noncompact if  $D = 1$  (the nonramified case), and in this case we have  $\mathcal{X}(1, N) = \Gamma_0(N) \backslash \mathbb{H}$ . If  $D > 1$ , then  $\mathcal{X}(D, N)$  is a compact surface. As we mentioned in section 1, we restrict to the level one case. The curves  $\mathcal{X}(D, 1)$ , for  $D > 1$ , can be viewed as the compact analogues of the modular surface  $\mathcal{X}(1, 1) = \mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}$ . The notion of CM points is now analogous to the case of the modular surface (see [1, Def. 6.4]) and we use the crucial fact that the class number  $h(D, d)$  is always finite.

**2.2. Hecke operators.** When the group  $\Gamma$  is arithmetic, there exists a large commutative algebra (Hecke algebra) of invariant, self-adjoint operators, the Hecke operators. They are defined by averaging over the orbits  $R(1) \backslash R(n)$ : for every  $n \geq 1$ , the  $n$ -th Hecke operator  $T_n : L^2(\mathcal{X}(D, 1)) \rightarrow L^2(\mathcal{X}(D, 1))$  is defined by

$$(2.6) \quad T_n f(z) = \sum_{x \in R(1) \backslash R(n)} f(\phi(x)z).$$

Since we are working in the level one case, for every  $n \geq 1$  the Hecke operator  $T_n$  is self-adjoint and commutes with the Laplacian  $-\Delta$ . Hence, we can choose a basis of eigenfunctions  $u_j$  which

simultaneously diagonalizes the Hecke operators and  $-\Delta$ . Such an eigenfunction is called a Hecke-Maass form. We always assume that  $u_j$  are  $L^2$ -normalized.

**2.3. Remarks on the explicit Jacquet-Langlands correspondence.** Jacquet and Langlands proved that there exists a nice, bijective correspondence between the nontrivial automorphic forms on the multiplicative group of a division quaternion algebra and certain cusp forms on  $GL(2)$ . The explicit Jacquet-Langlands correspondence was worked out by Hejhal [16], Koyama [24] (in view of an application to the Prime geodesic theorem), Bolte and Johansson [6] and Strömbergsson [38], and is given by the theta map (see [38]). It was further investigated by Risager [36], Arenas [2] and Blackman and Lemurell [5].

We refer to the explicit Jacquet-Langlands correspondence in various places; it may be possible that one can improve on some of our estimates by referring further to the explicit correspondence between  $\Gamma(D, 1)$  and  $\Gamma_0(D)$ . First, following the method of Iwaniec and Sarnak it may be possible to improve bound (2.9) on average. Secondly, Nelson [31] used the theta correspondence to study the quantum variance for cocompact groups. In principle, one may be able to use this to improve on Biro's result for  $\Gamma(D, N)$ . However, Biro's result is more spectral in nature than arithmetic whereas the explicit study of the quantum variance in [31] is quite arithmetic.

**2.4. Weyl's law and sup-norm estimates.** We will apply estimates for the growth of Maass forms in the spectral limit. Weyl's law describes the asymptotics of the number of Laplacian eigenvalues up to a fixed height; more precisely for any cocompact Fuchsian group  $\Gamma$  we have the asymptotic

$$(2.7) \quad \sum_{\lambda_j \leq T^2} 1 \sim \sum_{|t_j| \leq T} 1 \sim \frac{\text{vol}(\Gamma \backslash \mathbb{H})}{4\pi} T^2,$$

see [22, Corollary 11.2]. The following estimate is the local version of Weyl's law for  $L^2(\Gamma \backslash \mathbb{H})$ .

**Theorem 2.1** (Local Weyl's law). *For every  $z$ , as  $T \rightarrow \infty$ ,*

$$\sum_{|t_j| < T} |u_j(z)|^2 \sim cT^2,$$

where the constant  $c = c(z)$  depends only on the number of elements of  $\Gamma$  fixing  $z$ .

See [35, Lemma 2.3] for a proof of this result. We emphasize that if  $z$  remains in a compact set of  $\mathbb{H}$  the constant  $c(z)$  remains uniformly bounded. Thus, for compact surfaces we trivially have the uniform upper bound

$$(2.8) \quad \left\| \sum_{T \leq |t_j| \leq 2T} |u_j(z)|^2 \right\|_{\infty} \ll T^2$$

This follows also from Bessel's inequality [22, Proposition 7.2] and the fact that the height function  $y_{\Gamma}(z)$  is uniformly bounded for  $\Gamma$  cocompact. On the other hand, for sup norm estimates of Hecke-Maass forms  $u_j(z)$  on compact arithmetic surfaces we have the following deep subconvexity bound of Iwaniec and Sarnak [23]

$$(2.9) \quad \|u_j(z)\|_{\infty} \ll |t_j|^{5/12+\epsilon}$$

as  $\lambda_j \rightarrow \infty$ . Conjecturally, the sup-norm of the Maass-Hecke cusp form is expected to be bounded by  $t_j^{\epsilon}$ . Assuming that one would be able to control the dependence on  $D$  (i.e. on  $\Gamma$ ) in all the  $O$ -estimates for the local and the discrete average, it would be interesting to study the dependence of bound (1.7) on  $D \rightarrow \infty$ .

**2.5. The equidistribution of CM points, Weyl sums and bounds for associated  $L$ -functions.**

Equidistribution of CM points has been extensively studied by various authors. Duke proved his celebrated result about Heegner points on the modular surface using estimates for the Fourier coefficients of Maass forms (see also [27]). Michel [26] proved the equidistribution result on small Galois orbits for the case of Shimura curves associated to definite quaternion algebras over  $\mathbb{Q}$  and Harcos-Michel [15] worked the indefinite case. Zhang [40] generalized the equidistribution result for CM points on Shimura varieties.

**Theorem 2.2.** [11, Theorem 1], [27, Theorem 9] *Let  $\Gamma = \Gamma(D, N)$  be the Fuchsian group as above and let  $h(D, N, d)$  denote the number of CM points on  $\mathcal{X}(D, N)$  of fundamental discriminant  $d < 0$ . Then, for any compactly supported  $f$  on  $\mathcal{X}(D, N)$  we have as  $d \rightarrow -\infty$ ,*

$$(2.10) \quad \frac{1}{h(D, N, d)} \sum_{z \in \Lambda_d} f(z) \rightarrow \frac{1}{\text{vol}(\mathcal{X}(D, N))} \int_{\mathcal{X}(D, N)} f(z) d\mu(z).$$

*That means the CM points equidistribute on the Riemann surface  $\mathcal{X}(D, N)$  with respect to the hyperbolic measure.*

To understand the discrete average of the error term for the lattice counting problem we will apply the pre-trace formula. This allows us to write the average of the error over the CM points in relation to the average of Maass forms over the CM points, the so-called Weyl sums:

$$W(d, t_j) := \sum_{z \in \Lambda_d} u_j(z).$$

The effectivity in bound (2.10) is measured by estimates for the corresponding Weyl sums and the explicit Jacquet-Langlands correspondence.

We can bound the sum  $W(d, t_j)$  referring to a Waldspurger-Zhang type formula. We refer to [14, Lecture 2], [15, Section 6], [27, Section 2] for extended comments on this formula. Here, we focus on a bound on the  $d$ -aspect. By [15, Section 6] for  $|t_j| \rightarrow \infty$  we deduce the bound

$$(2.11) \quad \frac{W(d, t_j)}{h(D, d)} \ll |t_j|^A |d|^{-\frac{1}{28}},$$

The class number satisfies the classical bound  $|d|^{1/2-\epsilon} \ll h(D, d) \ll |d|^{1/2+\epsilon}$  (see [1, Proposition 3.55, Theorem 4.19]). The Weyl sum is related to central values of Maass forms through a general Waldspurger-Zhang formula in [15, page 647]:

$$(2.12) \quad |W(d, t_j)|^2 = c_{d, u_j} \frac{\sqrt{|d|} L(u_j \times \chi_d, 1/2) L(u_j, 1/2)}{L(\text{Sym}^2 u_j, 1)}$$

where  $c_{d, u_j} \ll (1 + |dt_j|)^\epsilon$ . Convexity bounds for the  $L$ -factors imply  $A = 1/2$  in (2.11). Further, we can obtain the following estimate:

**Proposition 2.3.** *For the Weyl sums we have the upper bound*

$$(2.13) \quad \sum_{T \leq t_j \leq 2T} \frac{|W(d, t_j)|^2}{h(D, d)^2} \ll |d|^{-\frac{1}{6} + \epsilon} T^{2+\epsilon}.$$

Using Ivić [21, page 455], Young [39, Theorem 1.1], Hoffstein-Lockhart [18, Theorem 0.1] and Jacquet-Langlands correspondence, the estimate follows applying the Hölder inequality, see [34, Lemma 2.2]. The conjectural Lindelöf estimate

$$L(u_j \times \chi_d, 1/2) \ll (1 + |dt_j|)^\epsilon$$

implies the exponent  $|d|^{-\frac{1}{2} + \epsilon}$  in (2.13).

**2.6. The local average of the error term for general cofinite groups.** We now explain the main idea relating the discrete average of the error term  $E(X; z, z)$  over CM points with the local average over a compact set. Roughly speaking, in the proof of the Theorem 1.2 we bound the discrete average of the error over the CM points by two pieces:

$$(2.14) \quad \frac{1}{h(D, d)} \sum_{z \in \Lambda_d} f(z) E(X; z, z) \ll E_{local}(X) + E(X),$$

where the  $E_{local}$  is the average of the error over compact sets of  $\Gamma \backslash \mathbb{H}$ :

$$E_{local}(X) := \int_{\Gamma \backslash \mathbb{H}} f(z) E(X; z, z) d\mu(z).$$

To prove our result we need bounds for both summands of (2.14), which we have to treat them separately. The tools we discussed in subsections 2.4 and 2.5 will be useful for dealing with the term  $E(X)$ , but the term  $E_{local}(X)$  has to be controlled differently.

In the case of the modular group  $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ , the local and discrete averages were studied separately in the independent works [33], [34]. The treatment of the local average uses the classical result of Luo-Sarnak [25] for the Quantum Unique Ergodicity of Maass cusp forms on average and can be applied for similar groups (arithmetic groups with cusps). Instead of using any strong arithmetic QUE information, Biro recently proved the following local average for general surfaces of finite area using a modified Selberg trace formula. For the case of compact arithmetic surfaces we apply his result.

**Theorem 2.4.** [4, Theorem 1.1] *Let  $\Gamma$  be a cofinite Fuchsian group and let  $f$  be a non-negative smooth compactly supported function on  $\Gamma \backslash \mathbb{H}$ . Then, as  $X \rightarrow \infty$ , the local average of the error term satisfies the bound*

$$(2.15) \quad \int_{\Gamma \backslash \mathbb{H}} f(z) E(X; z, z) d\mu(z) = O_{f, \Gamma, \epsilon}(X^{5/8+\epsilon}).$$

Biro was able to deduce his result by using his modification of the Selberg trace formula from [3]. In this work, instead of estimating the integral of the automorphic kernel  $K(z, w)$  over the diagonal, Biro worked with the integral of the automorphic kernel against some automorphic function  $u$  of eigenvalue  $\lambda$ . On this way, he deduced an identity between the geometric and the spectral side, where in the spectral side of the new trace formula he appeared the integrals

$$\int_{\Gamma \backslash \mathbb{H}} |u_j(z)|^2 u(z) d\mu(z).$$

in the place of the usual  $L^2$ -norms of  $u_j$  appearing in Selberg's formula. In the geometric side, he appeared weighted orbital integrals in the place of regular orbital integrals for the elliptic and hyperbolic terms (where the weight function depends on periods of  $u$ ). The appearance of the period integrals of  $u$  on the geometric side is the key ingredient that imply the improvement on the error term in (2.15).

**2.7. Bounds for inner products of eigenfunctions.** We refer to the following uniform upper bound for the inner product of two eigenfunctions. The proof given in [34, Theorem 5.1] relies on bounds for weight  $k$  eigenfunctions of  $\Gamma$ . Although it holds for a general cofinite group  $\Gamma$ , we state its simplified version only for our special case of a cocompact group.

**Proposition 2.5.** *For  $f$  a smooth function on the compact surface  $\Gamma \backslash \mathbb{H}$ ,  $u_m, u_k$  two Maass forms and  $a > 0$  we have the following bound:*

$$(2.16) \quad \langle f |u_m|^2, u_k \rangle \ll_{a, f} \left( \frac{1 + |t_m|}{1 + |t_k|} \right)^a \|u_m\|_\infty \|u_m\|_2 \|u_k\|_2.$$

## 3. PROOF OF THE MAIN THEOREM

**3.1. Local average of a Maass form over CM points.** Before proceeding to the proof of our main result, we start with a proposition for the average of Maass forms multiplied with  $f$  over Heegner points. Set

$$(3.1) \quad G(d, t_j) = \frac{1}{h(D, d)} \sum_{z \in \Lambda_d} f(z) |u_j(z)|^2 - \frac{1}{\text{vol}(\mathcal{X}(D, 1))} \int_{\mathcal{X}(D, 1)} f(z) |u_j(z)|^2 d\mu(z).$$

We have the following estimate.

**Proposition 3.1.** *Let  $f$  be a smooth function on  $\mathcal{X}(D, 1)$  and  $u_j(z)$  a Maass form on  $\mathcal{X}(D, 1)$  with eigenvalue  $\lambda_j = 1/4 + t_j^2 > 1/4$ . Then*

$$G(d, t_j) = O_{f, \Gamma, \epsilon} (\|u_j\|_\infty |d|^{-1/12+\epsilon} |t_j|^{1+\epsilon}).$$

*Proof.* Applying the spectral expansion to  $F(z) = f(z) |u_j(z)|^2$  and averaging over the CM points we conclude

$$G(d, t_j) = \sum_{\lambda_k \leq 1/4} \langle F, u_k \rangle \frac{W(d, t_k)}{h(D, d)} + \sum_{t_k \in \mathbb{R} - \{0\}} \langle F, u_k \rangle \frac{W(d, t_k)}{h(D, d)},$$

where  $\langle \cdot, \cdot \rangle$  denotes the Petersson inner product in  $L^2(\mathcal{X}(D, 1))$ . For a small eigenvalue  $\lambda_k \leq 1/4$  we have

$$W(d, t_k) \ll_\Gamma h(D, d) \implies \langle F, u_k \rangle \frac{W(d, t_k)}{h(D, d)} \ll_\Gamma \|f\|_\infty.$$

We conclude the contribution of the small eigenvalues is  $O_{f, \Gamma}(1)$ . For the big eigenvalues  $\lambda_k \geq 1/4$ , Proposition 2.3 and Cauchy-Schwarz in the interval  $[T, 2T]$  imply

$$(3.2) \quad \sum_{T \leq t_k \leq 2T} \langle F, u_k \rangle \frac{W(d, t_k)}{h(D, d)} \ll \left( \sum_{T \leq t_k \leq 2T} |\langle F, u_k \rangle|^2 \right)^{1/2} |d|^{-\frac{1}{12}+\epsilon} (1+T)^{1+\epsilon}.$$

Applying the estimate of Proposition 2.5 for the sum over  $t_k \in [T, 2T]$ , since all Maass forms are  $L^2$ -normalized we get the left side of (3.3) is bounded by

$$\ll_{a, f} \|u_j\|_\infty (1 + |t_j|)^a |d|^{-\frac{1}{12}+\epsilon} (1+T)^{2-a+\epsilon}$$

for any  $a > 0$ . On the other hand, using Bessel's inequality we get

$$(3.3) \quad \left( \sum_{T \leq t_k \leq 2T} |\langle f(z) |u_j(z)|^2, u_k \rangle|^2 \right)^{1/2} \ll \|f(z) |u_j(z)|^2\|_2 \ll_f \|u_j\|_\infty \|u_j\|_2 = \|u_j\|_\infty$$

uniformly for any  $T$ . Balancing between the two expressions, for any  $A \gg 1$  sufficient large and fixed we get

$$\sum_{t_k \in \mathbb{R} - \{0\}} \langle F, u_k \rangle \frac{W(d, t_k)}{h(D, d)} \ll_{f, \Gamma, \epsilon} \|u_j\|_\infty |d|^{-\frac{1}{12}+\epsilon} (A^{1+\epsilon} + (1 + |t_j|)^a A^{2-a+\epsilon}).$$

Picking  $A \asymp t_j^{\frac{a}{a-1}}$  with  $a$  large enough we deduce the statement.  $\square$

**3.2. Proof of Theorem 1.2.** We can now give the proof of our theorem. By a standard argument, we need to apply the pre-trace formula to suitable chosen kernels. We refer to [7, Section 2], [34, Section 6] for more details on the calculations. For the proof we will refer to some standard estimates for the Selberg/Harish-Chandra transforms of the characteristic kernel  $k(u) = \chi_{[0, (X-2)/4]}(u)$  and smooth approximations of  $k(u)$  (see Proposition 3.2).

*Proof.* Define  $R = \cosh^{-1}(X/2)$ , denote by  $u(z, w)$  the standard point-pair invariant function

$$(3.4) \quad u(z, w) = \frac{|z - w|^2}{4\Im(z)\Im(w)},$$

and let  $k(u)$  be the kernel

$$(3.5) \quad k(u) = \chi_{[0, (X-2)/4]}(u),$$

where by  $\chi_{[A, B]}(u)$  we denote the characteristic function of the interval  $[A, B]$ . Recall that  $\rho(z, w) \leq R$  if and only if  $4u(z, w) \leq X - 2$ , and thus it is straightforward to see that the automorphic kernel of  $k(u)$  satisfies

$$(3.6) \quad N(X, z, z) = K(z, z) := \sum_{\gamma \in \Gamma} k(u(\gamma z, z)).$$

The first problem arising in the study of the hyperbolic counting problem is that  $k(u)$  is not an admissible kernel for the pre-trace formula; this can be checked using the asymptotic behaviour of its Selberg/Harish-Chandra transform  $h(t)$  of  $k$  (see [22, eq. (1.62)] for the definition of the transform), and one has to approximate  $k$  by some suitable test functions.

For this reason, for  $\delta > 0$  small (that will be specified later), we define the smoothed kernel  $k_\delta(u)$  by

$$(3.7) \quad k_\delta(u) = \frac{1}{4\pi \sinh^2(\delta/2)} \chi_{[0, (\cosh \delta - 1)/2]}(u).$$

We also define the hyperbolic convolution of two functions by

$$(3.8) \quad k_1 * k_2(u(z, w)) = \int_{\mathbb{H}} k_1(u(z, v)) k_2(u(v, w)) d\mu(v),$$

and the kernels  $k_\pm(u)$  defined by

$$(3.9) \quad k_\pm(u) = (\chi_{[0, (\cosh(R \pm \delta) - 1)/2]} * k_\delta)(u)$$

If  $h_i(t)$  stands for the S/H-C transform of  $k_i(u)$ , then the S/H-C transform of the hyperbolic convolution  $k_1 * k_2$  is given the product  $h_1(t)h_2(t)$  (see [7, pages 322-323]). Using this observation we can compute the S/H-C transform of kernels  $k_\pm(u)$ .

**Proposition 3.2.** *With notation as above, the S/H-C transforms  $h_\pm(t)$  of the kernels  $k_\pm(u)$  satisfy:*

a) *If  $\lambda_j < 1/4$ , i.e.  $s_j = 1/2 + it_j$  with  $t_j \notin \mathbb{R}$  then*

$$(3.10) \quad h_\pm(t_j) = \sqrt{\pi} \frac{\Gamma(|t_j|)}{\Gamma(|t_j| + 3/2)} X^{1/2 + |t_j|} + O(X\delta + X^{1/2}).$$

b) *If  $\lambda_j = 1/4$ , i.e.  $t_j = 0$  then*

$$(3.11) \quad h_\pm(0) = O(X^{1/2} \log X).$$

c) *If  $\lambda_j > 1/4$ , i.e.  $t_j \in \mathbb{R} - \{0\}$  then*

$$(3.12) \quad h_\pm(t_j) = O\left(\frac{X^{1/2}}{|t_j|^{3/2}} \min\left\{1, \frac{1}{(\delta|t_j|)^{3/2}}\right\}\right).$$

*For  $t$  real we also have the uniform bound  $h_\pm(t_j) = O(X^{1/2} \log X)$ .*

*Proof.* Part a) follows from [7, eq.(2.14), p. 323] and part b) from [7, Lemma 2.4, (d)]. Part c) is calculated in [33, eq. (5.5), (5.10)], [34, eq. (6.3), (6.4)].  $\square$

Since  $k_- \leq k \leq k_+$  [34, page 15], summing over the group we deduce the inequalities

$$K^-(z, z) \leq N(X, z, z) \leq K^+(z, z),$$



hence

$$(3.13) \quad \sum_{z \in \Lambda_d} f(z) E(X; z, z) \ll \min_{\pm} \left| \sum_{z \in \Lambda_d} f(z) (K^{\pm}(z, z) - M(X; z, z)) \right|.$$

The kernels  $k_{\pm}$  are admissible in the pre-trace formula, which implies the expansions

$$(3.14) \quad K^{\pm}(z, z) = \sum_{t_j} h^{\pm}(t_j) |u_j(z)|^2.$$

Since we have a finite number of eigenvalues  $0 \leq \lambda_j \leq 1/4$ , we conclude

$$(3.15) \quad \begin{aligned} & \frac{1}{h(D, d)} \sum_{z \in \Lambda_d} f(z) (K^{\pm}(z, z) - M(X; z, z)) \\ &= \sum_{t_j \in \mathbb{R}} \frac{h_{\pm}(t_j)}{h(D, d)} \sum_{z \in \Lambda_d} f(z) |u_j(z)|^2 + O_{f, \Gamma}(\delta X + X^{1/2} \log X). \end{aligned}$$

The last sum can be rewritten as

$$(3.16) \quad \begin{aligned} & \sum_{t_j \in \mathbb{R}} \frac{h_{\pm}(t_j)}{\text{vol}(\mathcal{X}(D, 1))} \int_{\mathcal{X}(D, 1)} f(z) |u_j(z)|^2 d + \sum_{t_j \in \mathbb{R}} h_{\pm}(t_j) G(d, t_j) \\ &+ O(\delta X + X^{1/2} \log X). \end{aligned}$$

By Theorem 2.4, the first sum in (3.16) is bounded by  $X^{5/8+\epsilon}$ . For the second sum, we first estimate  $G(d, t_j)$  in dyadic intervals as follows. For real  $t_j \in [T, 2T]$ , Proposition 3.1 and bound (2.9) implies

$$(3.17) \quad \sum_{T \leq t_j \leq 2T} G(d, t_j) \ll_{f, \Gamma, \epsilon} |d|^{-1/12+\epsilon} T^{41/12+\epsilon}.$$

Secondly, interchanging the sums and using inequality (2.8) we get

$$\sum_{t_j \in \mathbb{R}} \frac{1}{h(D, d)} \sum_{z \in \Lambda_d} f(z) |u_j(z)|^2 \ll_{f, \Gamma} T^2$$

hence

$$(3.18) \quad \sum_{T \leq t_j \leq 2T} G(d, t_j) \ll_{f, \Gamma} T^2.$$

Interpolation between estimates (3.17), (3.18) gives

$$(3.19) \quad \sum_{T \leq t_j \leq 2T} G(d, t_j) \ll_{f, \epsilon} T^{2-2a} |d|^{-\frac{a}{12}+\epsilon} T^{\frac{41}{12}a+\epsilon} = |d|^{-\frac{a}{12}+\epsilon} T^{2+\frac{17}{12}a+\epsilon}$$

for any  $a \in [0, 1]$ . We apply estimates for the S/H-C transform from part c) of Proposition 3.2 and estimate (3.19). We get

$$(3.20) \quad \begin{aligned} \sum_{t_j \in \mathbb{R}} h_{\pm}(t_j) G(d, t_j) &= \sum_{|t_j| < \delta^{-1}} h_{\pm}(t_j) G(d, t_j) + \sum_{|t_j| \geq \delta^1} h_{\pm}(t_j) G(d, t_j) \\ &\ll X^{1/2} |d|^{-a/12+\epsilon} \delta^{-1(1/2+17a/12+\epsilon)} \end{aligned}$$

under the condition  $a < 12/17$ . Thus, the discrete average of the error term is balanced by

$$(3.21) \quad O\left(X^{5/8+\epsilon} + X^{1/2} |d|^{-a/12+\epsilon} \delta^{-1(1/2+17a/12+\epsilon)} + \delta X\right).$$

Picking  $a$  close to  $12/17$  and balancing  $\delta = X^{-1/5} |d|^{-2/85}$  we complete the proof of the theorem.  $\square$

## REFERENCES

- [1] M. Alsina and P. Bayer. *Quaternion orders, quadratic forms, and Shimura curves*. CRM Monograph Series, 22. American Mathematical Society, Providence, RI, 2004. xvi+196 pp.
- [2] A. Arenas. *On the traces of Hecke operators*. J. Number Th. 100 (2003), no. 2, 307–312.
- [3] A. Biró. *On a generalization of the Selberg trace formula*. Acta Arith. 87 (1999), no. 4, 319–338.
- [4] A. Biró. *Local average of the hyperbolic circle problem for Fuchsian groups*. Mathematika 64 (2018), no. 1, 159–183.
- [5] T. R. Blackman and S. Lemurel. *Spectral correspondences for Maass waveforms on quaternion groups*. J. Number Theory 158 (2016), 1–22.
- [6] J. Bolte and S. Johansson. *A spectral correspondence for Maass waveforms*. Geom. Funct. Anal. 9 (1999), no. 6, 1128–1155.
- [7] F. Chamizo. *Some applications of large sieve in Riemann surfaces*. Acta Arith. 77 (1996), no. 4, 315–337.
- [8] D. Chatzakos.  *$\Omega$ -results for the hyperbolic lattice point problem*. Proc. Amer. Math. Soc. 145 (2017), no. 4, 1421–1437.
- [9] G. Cherubini and M. Risager. *On the variance of the error term in the hyperbolic circle problem*. (to appear in Revista Matemática Iberoamericana), arXiv:1512.04779 (2015).
- [10] J. Delsarte. *Sur le gitter fuchsien*. (French), C. R. Acad. Sci. Paris 214 (1942), 147–179.
- [11] W. Duke. *Hyperbolic distribution problems and half-integral weight Maass forms*. Invent. Math. 92 (1988), no. 1, 73–90.
- [12] A. Good. *Local analysis of Selberg’s trace formula*. Lecture Notes in Mathematics, 1040. Springer-Verlag, Berlin, 1983. i+128 pp.
- [13] P. Günther. *Gitterpunktprobleme in symmetrischen Riemannschen Rumen vom Rang 1*. Math. Nachr. 94 (1980), 5–27.
- [14] G. Harcos. *Equidistribution on the modular surface and L-functions*. Homogeneous flows, moduli spaces and arithmetic, 377–387, Clay Math. Proc., 10, Amer. Math. Soc., Providence, RI, 2010.
- [15] G. Harcos and Ph. Michel. *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points. II*. Invent. Math. 163 (2006), no. 3, 581–655.
- [16] D. Hejhal. *A classical approach to a well-known spectral correspondence on quaternion groups*. Number theory (New York, 198384), 127–196, Lecture Notes in Math., 1135, Springer, Berlin, 1985.
- [17] R. Hill and L. Parnowski. *The variance of the hyperbolic lattice point counting function*. Russ. J. Math. Phys. 12 (2005), no. 4, 472–482.
- [18] J. Hoffstein and P. Lockhart. *Coefficients of Maass forms and the Siegel zero*. With an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman. Ann. of Math. (2) 140 (1994), no. 1, 161–181.
- [19] H. Huber. *Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I*. Comment. Math. Helv. 30 (1956), 20–62.
- [20] H. Huber. *Ein Gitterpunktproblem in der hyperbolischen Ebene*. J. Reine Angew. Math. 496 (1998), 15–53.
- [21] A. Ivić. *On sums of Hecke series in short intervals*. J. Thor. Nombres Bordeaux 13 (2001), no. 2, 453–468.
- [22] H. Iwaniec. *Spectral methods of automorphic forms*. Second edition. Graduate Studies in Mathematics, 53. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002. xii+220 pp.
- [23] H. Iwaniec and P. Sarnak.  *$L^\infty$  norms of eigenfunctions of arithmetic surfaces*. Ann. of Math. (2) 141 (1995), no. 2, 301–320.
- [24] S. Koyama. *Prime geodesic theorem for arithmetic compact surfaces*. Int. Math. Res. Not. (1998), no. 8, 383–388.
- [25] W. Luo and P. Sarnak, *Quantum ergodicity of Eigenfunctions on  $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}^2$* , Inst. Hautes Études Sci. Publ. Math. **81** (1995), 207–237.
- [26] Ph. Michel. *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*. Ann. of Math. (2) 160 (2004), no. 1, 185–236.
- [27] P. Michel and A. Venkatesh. *Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik*. In International Congress of Mathematicians. Vol. II, pages 421–457. Eur. Math. Soc., Zürich, 2006.
- [28] P. Michel and A. Venkatesh. *The subconvexity problem for  $GL_2$* . Publ. Math. Inst. Hautes tudes Sci. No. 111 (2010), 171–271.
- [29] P. Nelson, *Quantum variance on quaternion algebras I*, arXiv:1601.02526
- [30] P. Nelson, *Quantum variance on quaternion algebras II*, arXiv:1702.02669
- [31] P. Nelson, *Quantum variance on quaternion algebras III*, arXiv:1903.08686
- [32] S. J. Patterson. *A lattice-point problem in hyperbolic space*. Mathematika 22 (1975), no. 1, 81–88.
- [33] Y. Petridis and M. Risager. *Local average in hyperbolic lattice point counting, with an Appendix by Niko Laaksonen*. Math. Z. 285 (2017), no. 3–4, 1319–1344.
- [34] Y. Petridis and M. Risager. *Averaging over Heegner points in the hyperbolic circle problem*. Int. Math. Res. Not. 2018, no. 16, 4942–4968.
- [35] R. Phillips and Z. Rudnick. *The circle problem in the hyperbolic plane*. J. Funct. Anal. 121 (1994), no. 1, 78–116.
- [36] M. S. Risager. *Asymptotic densities of Maass newforms*. J. Number Theory 109 (2004), no. 1, 96–119.
- [37] A. Selberg. *Equidistribution in discrete groups and the spectral theory of automorphic forms*. <http://publications.ias.edu/selberg/section/2491>

- [38] A. Strömbergsson. *Some remarks on a spectral correspondence for Maass waveforms*. Int. Math. Res. Not. (2001), no. 10, 505–517.
- [39] M. Young. *Weyl-type hybrid subconvexity bounds for twisted  $L$ -functions and Heegner points on shrinking sets*. J. Eur. Math. Soc. 19 (2017), no. 5, 1545–1576.
- [40] S.-W. Zhang. *Equidistribution of CM-Points on quaternion Shimura varieties* Int. Math. Res. Not. (2005), no. 59, 3657–3689.

DEPARTAMENT DE MATEMÀTIQUES, ESCOLA POLITÈCNICA SUPERIOR D'ENGINYERIA DE MANRESA, UNIVERSITAT POLITÈCNICA DE CATALUNYA

*E-mail address:* `montserrat.alsina@upc.edu`

UNIVERSITÉ DE LILLE 1 SCIENCES ET TECHNOLOGIES AND CENTRE EUROPEEN POUR LES MATHÉMATIQUES, LA PHYSIQUE ET LEURS INTERACTIONS (CEMPI), CITÉ SCIENTIFIQUE, 59655 VILLENEUVE D'ASCQ CÉDEX, FRANCE

*E-mail address:* `dimitrios.chatzakos@univ-lille.fr`