



HAL
open science

Extreme values of geodesic periods on arithmetic hyperbolic surfaces

Bart Michels

► **To cite this version:**

Bart Michels. Extreme values of geodesic periods on arithmetic hyperbolic surfaces. 2020. hal-02460213v1

HAL Id: hal-02460213

<https://hal.science/hal-02460213v1>

Preprint submitted on 29 Jan 2020 (v1), last revised 12 Feb 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EXTREME VALUES OF GEODESIC PERIODS ON ARITHMETIC HYPERBOLIC SURFACES

BART MICHELS

ABSTRACT. Given a closed geodesic on a compact arithmetic hyperbolic surface, we show the existence of a sequence of Laplacian eigenfunctions whose integrals along the geodesic exhibit nontrivial growth. Via Waldspurger’s formula we deduce a lower bound for central values of Rankin–Selberg L -functions of Maass forms times theta series associated to real quadratic fields.

1. INTRODUCTION

Let X be a compact arithmetic hyperbolic surface and $-\Delta$ its Laplace operator, a self-adjoint operator on $L^2(X)$. The Hilbert space $L^2(X)$ admits an orthonormal basis $(\phi_j)_{j \geq 0}$ of real-valued smooth simultaneous eigenfunctions for $-\Delta$ and the Hecke algebra, ordered by nondecreasing Laplace eigenvalue $\lambda_j \geq 0$. It is known since Waldspurger [38] (see also [30]) that the integrals of the ϕ_j over special orbits on X are related to L -functions, making those *compact periods* an interesting object of study. When $z \in X$ is a CM-point, associated to an imaginary quadratic field, it was shown in [19] that the sequence of point evaluations $|\phi_j(z)|$ is unbounded: there exists a subsequence of (ϕ_j) on which

$$(1) \quad |\phi_j(z)| \gg \sqrt{\log \log \lambda_j}.$$

In [22], this was strengthened to the existence of a subsequence on which

$$(2) \quad |\phi_j(z)| \gg \exp\left(C \sqrt{\frac{\log \lambda_j}{\log \log \lambda_j}}\right)$$

for all $C < 1$ (and in fact, with the constant C replaced by 1 plus an error term, as in Theorem 1.1 below). In this context, the period formula relating the $\phi_j(z)$ to central values of L -functions is proven in [43]. To put these lower bounds in perspective, the local Weyl law [4, 17] implies the convexity bound $|\phi_j(z)| \ll \lambda_j^{1/4}$, and implies that $|\phi_j(z)| \asymp 1$ ‘on average’. Note that (1) and (2) imply a lower bound for the sup norms $\|\phi_j\|_\infty$. The *sup norm conjecture* of Iwaniec and Sarnak [19] asserts that $\|\phi_j\|_\infty \ll_\epsilon \lambda_j^\epsilon$ as $\lambda_j \rightarrow \infty$.

In this article, we prove a lower bound for integrals of eigenfunctions along closed geodesics, which are associated to *real* quadratic fields. The lower bound is of a quality similar to (2), relative to the average value of such geodesic periods. To quantify what ‘average’ means, let $\ell \subset X$ be a closed geodesic, and denote by $\mathcal{P}_\ell(\phi_j)$ the integral of ϕ_j along ℓ . In [41] it is shown that

$$\sum_{\lambda_j \leq \lambda} \mathcal{P}_\ell(\phi_j)^2 = C_\ell \lambda^{1/2} + O(1) \quad (\lambda \rightarrow \infty),$$

for some $C_\ell > 0$. This implies that $|\mathcal{P}_\ell(\phi_j)| \ll 1$, which is the convexity bound in this setting. By the Weyl law, the above sum consists of $\asymp \lambda$ terms, so that $\mathcal{P}_\ell(\phi_j)^2 \asymp \lambda^{-1/2}$ ‘on average’. In relation to a Waldspurger-type formula, the Lindelöf hypothesis for certain L -functions leads to the conjecture that $\lambda_j^{1/4} |\mathcal{P}_\ell(\phi_j)| \ll_\epsilon \lambda_j^\epsilon$ [26]. A geodesic period is ‘large’ when the normalized period $\lambda_j^{1/4} |\mathcal{P}_\ell(\phi_j)|$ is substantially larger than 1. We can now state our main theorem:

Theorem 1.1. *Let R be an Eichler order of square-free level in an indefinite quaternion division algebra over \mathbb{Q} . Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be the corresponding co-compact arithmetic lattice. Let (ϕ_j) be an orthonormal basis of $L^2(\Gamma \backslash \mathfrak{h})$ consisting of Laplace–Hecke eigenfunctions (see §2.6) with Laplacian eigenvalues $\lambda_j > 0$. Let*

We remark that the lower bound in (4) is of a quality similar to up-to-date lower bounds for extreme central values in various families of L -functions, obtained using the resonance method of Soundararajan [33] and Hilberdink [16]. Typically, one considers the Riemann zeta function on the critical line, the central value of Dirichlet L -functions of characters to large moduli or the central value of L -functions of modular forms of large weight. We refer to [6, 7, 10] for an account of recent results.

In [13] the authors explore how far such results are from being optimal, obtaining conjectures based on random models for L -functions: when \mathcal{F} is a suitable family of L -functions, there exists $B > 0$ such that the maximum of $|L(1/2)|$ when L runs through the L -functions in \mathcal{F} with analytic conductor at most D , is

$$\exp\left((B + o(1))\sqrt{\log D \log \log D}\right)$$

as $D \rightarrow \infty$. By Waldspurger's formula, this would imply that (3) is true with the exponent in the RHS replaced by $B\sqrt{\log \lambda \log \log \lambda}$ for some $B > 0$, if the max in the LHS runs over $\lambda_j \in [0, \lambda]$.

The proof of Theorem 1.1 uses the celebrated amplification method of Iwaniec and Sarnak [19], whose application to extreme values is reminiscent of the resonance method. The amplification method consists of comparing two trace formulas with an appropriate choice of adelic test functions. It exploits the presence of the Hecke operators, and the fact that many Hecke returns fix a given set (a CM-point or a closed geodesic). This is the approach in [19] and [22], which establish (1) and (2) respectively, and we follow it as well. We send the reader to §1.3 for an outline. The amplification method can be thought of as a way to quantify the phenomenon that eigenfunctions have strong concentration properties at points or geodesics that are fixed by many symmetries, a phenomenon that is also observed for zonal spherical harmonics and Gaussian beams on round spheres.

The method stands in contrast to another strategy to prove the existence of large values of compact periods, which uses functoriality and a vanishing property to show that a certain sparse subsequence of (ϕ_j) must attain large values on a distinguished set; see e.g. [27]. In some cases, the eigenfunctions one obtains by the amplification method are also related to functoriality (see [8, 23] for a discussion of the relation between the two methods). But on hyperbolic surfaces, that does not seem to be the case.

1.1. The random wave conjecture and large values. The fact that on arithmetic hyperbolic surfaces the sequence $\|\phi_j\|_\infty$ is unbounded, can, as remarked above, be explained through random models for L -functions, but also by random behavior of Laplacian eigenfunctions. In this subsection and the next, we explore how much growth can be explained by randomness of eigenfunctions, and we discuss some recent developments and related results.

Unless otherwise stated, let X be a compact Riemannian surface with negative sectional curvature. (Some of the results and conjectures stated will apply to arbitrary curvature or arbitrary dimension, but generality is not our objective.) Let (ϕ_j) be an orthonormal basis of $L^2(X)$ consisting of Laplacian eigenfunctions. (If X is a noncompact finite-volume quotient of \mathfrak{h} , we take it to be an orthonormal basis of the space of Maass cusp forms.) Because X is compact and of negative curvature, its geodesic flow is ergodic, and the *random wave conjecture* of Berry [5] predicts that Laplacian eigenfunctions of large eigenvalue should show

Gaussian random behavior. Berry’s conjecture has been tested numerically in [2, 3] for certain compact hyperbolic surfaces of genus 2, and in [15] for the modular surface $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathfrak{h}$ (which, although non-compact, still has ergodic geodesic flow). In this last article, the authors propose the following mathematical interpretation of Berry’s conjecture: Equip X with the normalized Riemannian probability measure $\mathrm{Vol}(X)^{-1}d\mathrm{Vol}$. Then the sequence of eigenfunctions (ϕ_j) , considered as random variables on X , converges weakly to the normal distribution with mean 0 and variance $\mathrm{Vol}(X)^{-1}$. The random wave conjecture has been the subject of recent work [1] where it has been interpreted in terms of Benjamini–Schramm convergence of manifolds. In [1, Theorem 4] is also shown that this version of the conjecture implies the quantum unique ergodicity conjecture (QUE) of Rudnick and Sarnak [27] which states that the probability measures $\phi_j^2 d\mathrm{Vol}$ converge weakly to the uniform measure $\mathrm{Vol}(X)^{-1}d\mathrm{Vol}$.

Even though Berry’s conjecture remains widely open in any interpretation, it is natural to wonder about the finer statistical properties of Laplacian eigenfunctions. For example, when the random wave conjecture is regarded as an analogue of the central limit theorem, one could ask what the analogue of the law of the iterated logarithm should be. That is, what can be said about the growth of the sup norm $\|\phi_j\|_\infty$ as $j \rightarrow \infty$? In order to obtain predictions for this type of questions, it is common to study *random wave models*, probability measures on a space of functions. Following [42, §6], we may model an eigenfunction ϕ of large eigenvalue λ by the random variable

$$\frac{1}{(\sum_j c_j^2)^{1/2}} \sum_j c_j \phi_j$$

where the c_j are i.i.d. standard Gaussians and the sums run over the set $\{j : \sqrt{\lambda_j} \in [\sqrt{\lambda}, \sqrt{\lambda} + 1]\}$. One thus expects the sup norm $\|\phi\|_\infty$ to generically have the same behavior as the sup norm of random Fourier series studied in [20, §6], [28, Chapter IV]. That is, that $\|\phi_j\|_\infty \asymp \sqrt{\log \lambda_j}$ for a subsequence of (ϕ_j) [15, 30]. This is compatible with the sup norm conjecture and with (1). As for (2), note that the RHS is eventually smaller than any power of λ_j , and eventually larger than any power of $\log \lambda_j$. Such large sup norms, while not in contradiction with the sup norm conjecture nor the random wave conjecture, should be considered exceptional and are presumably specific to the case of arithmetic manifolds: in both articles, one uses the recurrence properties of CM-points under Hecke operators. Considering these results, it is somewhat surprising that despite the presence of Hecke operators, numerical evidence indicates that the random wave conjecture does hold for arithmetic hyperbolic surfaces.

1.2. Restricted randomness and random restrictions. One can think of at least two ways to generalize questions about random behavior of the ϕ_j . On the one hand, one can fix a ‘thin’ set S , such as a finite set or a geodesic segment, equipped with its induced Riemannian metric and volume element μ_S , and ask about the value distribution of the restrictions $\phi_j|_S$. It is clear that any result must be very sensitive to the nature of the set S . We illustrate this with an example. Quantum (unique) ergodic restriction, Q(U)ER, is the question of whether a density one subsequence (resp. the full sequence) of the restrictions $\phi_j^2|_S$ equidistribute. It was first studied in [35]. When X is the modular surface, the sequence (ϕ_j) of Maass cusp forms can be chosen to consist of odd and even cusp forms. The odd ones

vanish on the split geodesic $S = [i, i\infty)$, and one expects QER to hold for the restrictions to S of the even Maass forms, but where the limiting measure is $2\mu_S$ instead of μ_S [39]. We mention also [40], where the analogue of QER for Eisenstein series is found to have a negative answer on certain divergent geodesics on the modular surface. In the other direction, QER does hold for generic geodesics on a compact hyperbolic surface [11, 36].

The other point of view consists of randomizing the thin set. Given an eigenfunction and a geodesic segment L , we may form the integral

$$\mathcal{P}_L(\phi_j) := \int_L \phi_j.$$

If we fix a real number $l > 0$, the set of geodesic segments of length l is parametrized by the unit tangent bundle S^1X . Every eigenfunction ϕ_j gives rise to a smooth function $\tilde{\phi}_j$ on S^1X , by integrating over the segment corresponding to a point of S^1X . It is known that the sequence $\|\tilde{\phi}_j\|_\infty$ is bounded as $\lambda_j \rightarrow \infty$ (by a constant depending on X and l) [9]. In particular, when ℓ is a closed geodesic, we recover the convexity bound $|\mathcal{P}_\ell(\phi_j)| \ll 1$, which is sharp when X is the round 2-sphere. Theorem 1.1 can be viewed as producing large values of the functions $\tilde{\phi}_j$ for suitable l .

We remark that, in order to obtain heuristics for large values of $\tilde{\phi}_j$, to apply random wave models as in §1.1 would be nonsensical, because Berry's conjecture is fundamentally a statement about the value distribution of eigenfunctions, and does not take into account the relative position of points. That is, the random wave models are not designed to be integrated over sets of measure zero. Instead, it would be interesting to develop a conjecture analogous to Berry's for the value distribution of $\tilde{\phi}_j$ as $\lambda_j \rightarrow \infty$. A related type of result is the central limit theorem for geodesic flows [25, 32]: for a fixed smooth function $\phi \in C^\infty(X)$, one determines the limiting distribution of the (normalized) integral of ϕ along a random geodesic segment of growing length.

1.3. Organization of the paper. The proof of Theorem 1.1 goes by a comparison of trace formulas. One is the Selberg trace formula; the other is a relative (pre)-trace formula involving geodesic periods. In §3, we derive both formulas and polish the first by identifying a main term and bounding the error term. Here, a novelty is the estimation of the contribution of hyperbolic classes in the Selberg trace formula, which is an unavoidable issue given the nature of the problem; see Remark 3.1. In the relative pre-trace formula, we similarly identify a main term, understanding which involves an arithmetic (§4.1) and an analytic question (§5). Bounding the error term leads to a Diophantine problem (§4.2) and the analytic problem of bounding orbital integrals (§6). In §7.1, we combine the work and consider an arbitrary amplifier, obtaining an asymptotic estimate for short spectral sums. In §7.2 we optimize the amplifier, proving Theorem 1.1. In §7.3 we adapt the method to prove Corollary 1.2.

1.4. Acknowledgements. I thank Farrell Brumley for giving me this problem, and for helpful suggestions. I also thank Nicolas Bergeron and Étienne Le Masson for useful conversations surrounding the topic of this paper.

2. NOTATION

2.1. Lie groups. Let $G = \mathrm{PSL}_2(\mathbb{R})$ and define the subgroups $K = \mathrm{PSO}_2(\mathbb{R})$, $A = \left\{ \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \right\}$ and $N = \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\}$. Fix parametrizations $a : \mathbb{R} \rightarrow A$, $k : \mathbb{R}/2\pi\mathbb{Z} \rightarrow K$ defined by

$$a(t) = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \quad k(\theta) = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}.$$

The action of G on the hyperbolic plane \mathfrak{h} induces a map $G \rightarrow \mathfrak{h} : g \mapsto gi$ and a diffeomorphism $N \times A \cong \mathfrak{h}$. Denote by dg be the (bi-invariant) Haar measure on $G \cong \mathfrak{h} \times K$ that is the product of the hyperbolic measure with the Haar measure of mass 1 on K . Fix any Riemannian metric on G , and denote the Riemannian distance on G by $d = d_G$.

We will often use the fact that any two Riemannian distances are locally equivalent, or more generally the following fact: When M is a smooth connected manifold, $N \subset M$ a connected submanifold, each equipped with a Riemannian metric, then for every compact subset $L \subset N$ and $x, y \in L$ we have $d_N(x, y) \asymp_L d_M(x, y)$. In particular, taking $M = N = G$ and pulling back the metric on G on the left or the right by $g \in G$, we have for every compact $L \subset G$ and $x, y \in L$ that $d(gx, gy) \asymp_{g,L} d(xg, yg) \asymp_{g,L} d(x, y)$. Taking $N = \mathrm{SL}_2(\mathbb{R})$ equipped with any Riemannian metric and $M = M_2(\mathbb{R})$ with the Euclidean metric w.r.t. the standard basis, we find that for $L \subset \mathrm{SL}_2(\mathbb{R})$ compact and $x, y \in L$, $d_{\mathrm{SL}_2}(x, y) \asymp_L \|x - y\|_2$. We also have $d_G(x, y) \asymp_L \min(d_{\mathrm{SL}_2}(x, y), d_{\mathrm{SL}_2}(x, -y))$.

2.2. Arithmetic quotients. Let B be a quaternion division algebra over \mathbb{Q} which is split at ∞ . That is, there exists an isomorphism of \mathbb{R} -algebras $\rho : B \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow{\sim} M_2(\mathbb{R})$. We view B as a subset of $B \otimes_{\mathbb{Q}} \mathbb{R}$ via the natural embedding. Denote the projection $\mathrm{GL}_2^+(\mathbb{R}) \rightarrow \mathrm{PSL}_2(\mathbb{R}) = G$ by $g \mapsto \bar{g}$, and denote $\bar{\rho}(g) = \overline{\rho(g)}$ for an element $g \in (B \otimes_{\mathbb{Q}} \mathbb{R})^+$ of positive reduced (quaternion) norm. We will not always distinguish between $\eta \in B^+$ and its image $\bar{\rho}(\eta) \in G$. Let $R \subset B$ be a \mathbb{Z} -order. For $n \in \mathbb{N}_{>0}$, denote by $R(n) \subset B^+$ the set of elements of reduced norm equal to n , denote $R^1 = R(1)$ and define $\Gamma = \bar{\rho}(R^1) \subset G$. It is well known that Γ is a lattice in G , and that the quotient $\Gamma \backslash \mathfrak{h}$ is compact (see e.g. [37, IV: Théorème 1.1.]). We call Γ an arithmetic lattice.

Let $\|\cdot\|$ be the operator norm on $M_2(\mathbb{R}) \cong B \otimes_{\mathbb{Q}} \mathbb{R}$. It is invariant under quaternion conjugation.

2.3. Closed geodesics. A *geodesic* in \mathfrak{h} is the image of a maximal geodesic curve. The image of A in \mathfrak{h} is the geodesic joining 0 with ∞ . Because G acts transitively on the unit tangent bundle of \mathfrak{h} , it acts transitively on geodesics. If L is the geodesic that is the image of gA in \mathfrak{h} , its set-wise stabilizer is the normalizer $N_G(gAg^{-1})$, in which gAg^{-1} has index equal to 2.

If $\Gamma \subset G$ is a discrete subgroup, a geodesic $L = gAi \subset \mathfrak{h}$ projects to a smooth curve in the Riemannian orbifold $\Gamma \backslash \mathfrak{h}$. It has a periodic image $\ell \subset \Gamma \backslash \mathfrak{h}$ precisely when there exists $\gamma \in (\Gamma \cap gAg^{-1}) - \{1\}$ which stabilizes L . We then call ℓ a closed geodesic and denote $\Gamma_L = \mathrm{Stab}_{\Gamma}(L) \cap gAg^{-1}$. It is a lattice in gAg^{-1} and we have $[\mathrm{Stab}_{\Gamma}(L) : \Gamma_L] \in \{1, 2\}$. When this index equals 1, ℓ is called a reciprocal geodesic; see [31]. When ℓ is a closed geodesic and $\phi \in C^\infty(\Gamma \backslash \mathfrak{h})$, define the period of ϕ along

ℓ as the line integral

$$(5) \quad \mathcal{P}_\ell(\phi) := \int_{\Gamma_L \backslash L} \phi.$$

When Γ is an arithmetic lattice as in §2.2, the closed geodesics can be characterized as follows. When $F \subset B$ is a real quadratic number field, the \mathbb{R} -algebra $F \otimes_{\mathbb{Q}} \mathbb{R} \subset B \otimes_{\mathbb{Q}} \mathbb{R}$ is isomorphic to $\mathbb{R} \times \mathbb{R}$, hence its image under ρ is conjugate in $M_2(\mathbb{R})$ to the algebra of diagonal matrices. Thus the group $\bar{\rho}((F \otimes_{\mathbb{Q}} \mathbb{R})^1)$ of matrices of determinant ± 1 equals gAg^{-1} for a unique $g \in G/N_G(A)$. One has that gA projects to a closed geodesic in $\Gamma \backslash \mathfrak{h}$ and every closed geodesic is obtained exactly once in this way: we obtain a bijection $F \mapsto L_F$ between real quadratic number fields inside B and geodesics in \mathfrak{h} that become closed in $\Gamma \backslash \mathfrak{h}$. (In particular, whether a geodesic becomes closed in $\Gamma \backslash \mathfrak{h}$ depends only on B and ρ .) It induces a bijection between real quadratic $F \subset B$ modulo conjugation by R^1 , and closed geodesics $\ell \subset \Gamma \backslash \mathfrak{h}$.

When F is such a real quadratic field, we write $R_F = R \cap F$. It is an order in F and one has that $\Gamma_{L_F} = \bar{\rho}(R_F^1)$, whereas $\text{Stab}_\Gamma(L_F)$ equals the normalizer $\bar{\rho}(N_{R^1}(F))$. More generally, one has $\text{Stab}_{B^+}(L_F) = N_{B^+}(F)$. Let $\omega \in B^\times$ be a Skolem–Noether element w.r.t. F . That is, conjugation by ω leaves F invariant and induces the non-trivial automorphism of F . Then $N_{B^+}(F) = F^+ \sqcup (\omega F)^+$.

Remark 2.1. Closed geodesics are naturally grouped in packets indexed by a class group, and the above bijection gives the geodesic corresponding to the identity of the group. When R is an Eichler order of square-free level, packets can be described without use of the adelic language as follows: Let $F \subset B$ be a real quadratic field, and I an invertible fractional R_F -ideal. The right R -ideal IR is principal and generated by an element $a \in B^+$. To I we associate the R^1 -conjugacy class of the field $a^{-1}Fa$ (i.e., a closed geodesic). The map obtained in this way factors through the narrow class group of R_F to an injective map.

Remark 2.2. When a closed geodesic is viewed as a subset of $\Gamma \backslash \mathfrak{h}$, it may seem natural to define $\mathcal{P}_\ell(\phi_j)$ instead as an integral over $\text{Stab}_\Gamma(L) \backslash L$, resulting in a period that is half as large when L is a reciprocal geodesic. However, the definition (5) is closer to the notion of an adelic period, and appears naturally in a period formula, which we now state.

Remark 2.3. Geodesic periods are related to central values of Rankin–Selberg L -functions, as follows: [24, Theorem 5.4.1] Let $N \geq 1$ be a square-free integer, and let F be a real quadratic field of square-free discriminant d_F coprime to N . Assume that the number of primes dividing N which are inert in F is even. Let B be a quaternion algebra over \mathbb{Q} which is ramified exactly at the primes dividing N which are inert in F . Fix an embedding $\iota : F \hookrightarrow B$. Let $R \subset B$ be an Eichler order of level N/Δ_B containing the full ring of integers $\iota(\mathcal{O}_F)$. (The freedom in the choice of the Eichler order R is explained in [24, Remark 5.3.3].) Let $\Gamma \subset \text{PSL}_2(\mathbb{R})$ be the corresponding lattice. Let h_F^+ be the narrow class number of F . To F corresponds a packet Λ_{d_F} consisting of h_F^+ closed geodesics in $\Gamma \backslash \mathfrak{h}$. Let χ_0 be the trivial character of the narrow class group of F and π_{χ_0} the associated automorphic representation of $\text{GL}_2(\mathbb{A}_{\mathbb{Q}})$ given by the theta correspondence. Let f be an even Maass newform of level $\Gamma_0(N)$. Let π_f be the automorphic representation of $\text{GL}_2(\mathbb{A}_{\mathbb{Q}})$ it generates. Let π_f^{JL} be the corresponding representation of $B^\times(\mathbb{A}_{\mathbb{Q}})$ given by the Jacquet–Langlands correspondence and take a newvector in π_f^{JL} , which corresponds to an

L^2 -normalized Hecke–Laplace eigenfunction ϕ on $\Gamma \backslash \mathfrak{h}$. Then

$$(6) \quad \frac{\Lambda(1/2, \pi_f \times \pi_{\chi_0})}{\Lambda(1, \pi_f, \text{Ad})} = 2d_F^{-1/2} \prod_{p|\Delta_B} \frac{p+1}{p-1} \cdot \left| \sum_{\ell \in \Lambda_{d_F}} \mathcal{P}_\ell(\phi) \right|^2,$$

where the Λ 's in the LHS are completed Rankin–Selberg L -functions.

Note that: (1) The assumption that there is an even number of prime divisors of N that are inert in F , implies that the ϵ -factor in the functional equation for $\Lambda(s, \pi_f \times \pi_{\chi_0})$ is 1, so that there is no forced vanishing of the central value; (2) If there is *at least one* such prime divisor, B is a division algebra; (3) If *all* prime divisors of N are inert in F , R is a maximal order. In Corollary 1.2 we impose all three conditions.

Remark 2.4. If in Theorem 1.1 we drop the requirement that the subsequence of the (ϕ_j) consists of Hecke eigenfunctions, the statement holds for an arbitrary order R . Indeed, one may reduce to the case of a maximal order, as follows: if R is any order, let R' be a maximal order containing R , and Γ' the corresponding lattice. When ϕ' is an L^2 -normalized eigenfunction for $\Gamma' \backslash \mathfrak{h}$, then $\phi := [\Gamma' : \Gamma]^{-1/2} \phi'$ is one for the cover $\Gamma \backslash \mathfrak{h}$, and ϕ and ϕ' have the same Laplacian eigenvalue. The periods $\mathcal{P}(\phi)$ and $\mathcal{P}'(\phi')$ w.r.t. Γ and Γ' are related by

$$\mathcal{P}(\phi) = [\Gamma' : \Gamma]^{-1/2} [\Gamma'_L : \Gamma_L] \mathcal{P}'(\phi').$$

Thus large geodesic periods on $\Gamma' \backslash \mathfrak{h}$ produce large geodesic periods on $\Gamma \backslash \mathfrak{h}$. The reason we do restrict to Eichler orders of square-free level, is that the Hecke theory for such orders is well documented.

2.4. Hecke operators. Assume from now on that R is an Eichler order of square-free level. Let Δ_R be its discriminant and $\Gamma = \bar{\rho}(R^1)$. For $n \in \mathbb{N}_{>0}$, the quotients $R(1) \backslash R(n)$ are finite and we define the Hecke operators T_n , which act on $L^2(\Gamma \backslash \mathfrak{h})$, by

$$(T_n f)(z) = \sum_{\eta \in R(1) \backslash R(n)} f(\eta z).$$

The T_n are self-adjoint operators and commute with the Laplacian Δ on smooth functions. Because R is an Eichler order of square-free level, when $(mn, \Delta_R) = 1$ we have the relations

$$(7) \quad T_m T_n = T_n T_m = \sum_{d|m, n} d \cdot T_{mn/d^2}$$

(see [12, §III.7]). Note that the Hecke operator T_n is often normalized by multiplying the above sum by a factor $1/\sqrt{n}$, which we don't do here.

2.5. Convolution operators. In order to extract short spectral sums from the spectral side of the pre-trace formula, we will make use of convolution operators that are approximate spectral projectors, and whose Harish-Chandra transform satisfies certain positivity properties. Their existence is guaranteed by Proposition 2.5.

For $s \in \mathbb{C}$, define the spherical function $\varphi_s \in C^\infty(K \backslash G / K)$ by

$$\varphi_s(g) = \int_K e^{(1/2+s)H(kg)} dk,$$

where $H : G \rightarrow \mathbb{R}$ is defined by requiring that $g \in Na(H(g))K$. Define the Harish-Chandra transform

$$\widehat{k}(s) = \int_G k(g)\varphi_s(g)dg,$$

which is an entire function of finite type whose type depends on the support of k .

Proposition 2.5. *There exists a family $(k_\nu)_{\nu \geq 0}$ of bi- K -invariant smooth functions on G satisfying:*

- (1) *There exists $R > 0$ such that k_ν is supported in the ball $B(e, R)$ for all ν ;*
- (2) *$\widehat{k}_\nu(s) \in \mathbb{R}_{\geq 0}$ when $s \in \mathbb{R} \cup i\mathbb{R}$;*
- (3) *$\widehat{k}_\nu(ir) \geq 1$ for $|r - \nu| \leq 1$;*
- (4) *$\widehat{k}_\nu(ir) \ll_N (1 + |\nu - r|)^{-N}$ uniformly for $\nu \geq 0$ and $r \in \mathbb{R}_{\geq 0} \cup [-i/2, i/2]$;*
- (5) *$k_\nu(e) \asymp \nu$ as $\nu \rightarrow \infty$.*

Proof. Take any real-valued $k \in C_0^\infty(K \backslash G/K)$. The Harish-Chandra transform $\widehat{k}(s)$ is clearly real when $s \in \mathbb{R}$, and in view of the functional equation $\varphi_{-s} = \varphi_s$, also when $s \in i\mathbb{R}$. If we choose k nonnegative and not identically zero, we have $\widehat{k}(s) > 0$ for $s \in \mathbb{R}$; in particular for $s = 0$. Let $k_{(1)} = k * k$. Then $\widehat{k}_{(1)}(s) = \widehat{k}(s)^2 \geq 0$ for $s \in \mathbb{R} \cup i\mathbb{R}$. There exists $\delta > 0$ such that $\widehat{k}_{(1)}(s) > \widehat{k}_{(1)}(0)/2$ for $|s| \leq \delta$. Let $\widehat{k}_{(2)}(s) = 2\widehat{k}_{(1)}(\delta s)/\widehat{k}_{(1)}(0)$. Then $\widehat{k}_{(2)}(s) \geq 1$ for $|s| \leq 1$. Let $k_{(2)}$ be its inverse Harish-Chandra transform. Define

$$\widehat{k}_\nu(s) = \widehat{k}_{(2)}(i\nu + s) + \widehat{k}_{(2)}(i\nu - s).$$

Then \widehat{k}_ν is entire, even, of exponential type (at most) the exponential type of $\widehat{k}_{(2)}$, so it is the Harish-Chandra transform of a $k_\nu \in C_0^\infty(K \backslash G/K)$ whose support is bounded independently of ν . Conditions 1. and 2. are now satisfied. When $|r - \nu| \leq 1$, we have

$$\begin{aligned} \widehat{k}_\nu(ir) &= \widehat{k}_{(2)}(i\nu + ir) + \widehat{k}_{(2)}(i\nu - ir) \\ &\geq 0 + 1, \end{aligned}$$

which implies condition 3. Being the Fourier transform of the Abel transform of $k_{(2)}$, $\widehat{k}_{(2)}$ is of rapid decay on vertical strips of \mathbb{C} . Thus when $\operatorname{Re}(r) \geq 0$ and $\operatorname{Im}(r)$ remains bounded,

$$\begin{aligned} \widehat{k}_\nu(ir) &\ll_N (1 + |\nu + r|)^{-N} + (1 + |\nu - r|)^{-N} \\ &\ll (1 + |\nu - r|)^{-N}, \end{aligned}$$

which is 4. For 5., using the inverse Harish-Chandra transform we have

$$(8) \quad k_\nu(g) = \int_0^\infty \widehat{k}_\nu(ir)\varphi_{ir}(g)\beta(r)dr,$$

where $\beta(r) = \frac{1}{2\pi} \tanh(\pi r) \cdot r$ is the Plancherel density. When $g = e$, $\varphi_{ir}(g) = 1$ so that

$$k_\nu(e) \geq \int_\nu^{\nu+1} \widehat{k}_\nu(ir)\beta(r)dr \gg \nu$$

on the one hand, and

$$\begin{aligned} k_\nu(e) &\ll \int_0^{2\nu} \widehat{k}_\nu(ir) \nu dr + \int_{2\nu}^\infty \frac{r}{(r-\nu)^3} dr \\ &\ll \nu \cdot \|\widehat{k}_\nu\|_{L^1(i\mathbb{R})} + \nu^{-1} \\ &\ll \nu \end{aligned}$$

on the other. \square

Throughout, we fix a family (k_ν) as in Proposition 2.5.

2.6. Eigenfunctions. Let $(\phi_j)_{j \geq 0}$ be an orthonormal basis of $L^2(\Gamma \backslash \mathfrak{h})$ consisting of simultaneous eigenfunctions for Δ and the Hecke operators T_n for $(n, \Delta_R) = 1$, ordered by nondecreasing Laplacian eigenvalue $\lambda_j \geq 0$:

$$(\Delta + \lambda_j)\phi_j = 0.$$

Write the n th Hecke eigenvalue of ϕ_j as $\widehat{T}_n(\phi_j)$. For each j , write $\lambda_j = \frac{1}{4} + \nu_j^2$, where $\nu_j \in \mathbb{R}_{\geq 0} \cup [0, i/2]$ is called the spectral parameter of ϕ_j , to be thought of as the frequency of the wavefunction ϕ_j . The automorphic kernel $K_\nu(x, y) = \sum_{\gamma \in \Gamma} k_\nu(x^{-1}\gamma y)$ acts as an integral operator on $L^2(\Gamma \backslash \mathfrak{h})$, and the K_ν -eigenvalue of ϕ_j is $\widehat{k}_\nu(i\nu_j)$.

3. TWO TRACE FORMULAS

As in §2, let $\Gamma \subset G$ be the lattice coming from an Eichler order R of square-free level in a quaternion division algebra B , $F \subset B$ a real quadratic field, $L = L_F \subset \mathfrak{h}$ the corresponding geodesic and ℓ the corresponding closed geodesic. Denote $\mathcal{P}(\phi_j) = \mathcal{P}_\ell(\phi_j)$. Let $g_0 A g_0^{-1} = A_L$ be the identity component of $\text{Stab}_G(L)$.

We start from the spectral expansion of the automorphization of k_ν :

$$\sum_j \widehat{k}_\nu(i\nu_j) \phi_j(x) \phi_j(y) = \sum_{\gamma \in \Gamma} k_\nu(x^{-1}\gamma y),$$

where the convergence is uniform for x and y in compact sets. Let $n \geq 1$ be an integer coprime to the discriminant Δ_R , and apply T_n to the x -variable to obtain

$$(9) \quad \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) \phi_j(x) \phi_j(y) = \sum_{\eta \in R(n)/\pm 1} k_\nu(x^{-1}\eta y).$$

Setting $x = y$ and integrating over $\Gamma \backslash \mathfrak{h}$ we obtain the amplified standard pre-trace formula:

$$(10) \quad \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) = \int_{\Gamma \backslash \mathfrak{h}} \sum_{\eta \in R(n)/\pm 1} k_\nu(x^{-1}\eta x) dx$$

On the other hand, to make periods appear, we may integrate (9) over $\ell \times \ell$ to get the amplified relative pre-trace formula:

$$(11) \quad \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) \mathcal{P}(\phi_j)^2 = \int_{\ell \times \ell} \sum_{\eta \in R(n)/\pm 1} k_\nu(x^{-1}\eta y) dx dy$$

We now identify a main term and an error term in both pre-trace formulas.

3.1. Standard pre-trace formula. Rewrite (10) as

$$(12) \quad \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) = |(R(n) \cap \mathbb{Q}) / \pm 1| \cdot \text{Vol}(\Gamma \backslash \mathfrak{h}) \cdot k_\nu(e) \\ + \sum_{\eta \in (R(n) - \mathbb{Q}) / \pm 1} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx$$

where $\mathcal{F} \subset \mathfrak{h}$ is a fundamental domain for the action of Γ . We want to bound the sum in the RHS. We follow [19, §2]. The argument changes slightly because our convolution operator k_ν is a different one: (1) The support of k_ν does not decrease with ν , which makes that the contribution of hyperbolic elements cannot be ignored. (2) k_ν is not necessarily nonnegative, so we cannot bound the contribution of a single η by the contribution of its conjugacy class.

Remark 3.1. The reason why we (have to) make this choice of convolution operator k_ν , is the following. Our ultimate goal is to prove an inequality of the form

$$(13) \quad \sum_j \widehat{k}_\nu(i\nu_j) w_j \mathcal{P}(\phi_j)^2 \geq C_\nu \cdot \sum_j \widehat{k}_\nu(i\nu_j) w_j$$

for suitable nonnegative weights w_j , which will imply that at least one eigenfunction ϕ_{j_0} satisfies the bound $\mathcal{P}(\phi_{j_0})^2 \geq C_\nu$. Here, we expect C_ν to be just slightly larger than ν^{-1} . In particular, it decreases with ν (in contrast to the situation in [19, 22]). We want to bound C_ν from below by $\nu_{j_0}^{-1}$. If ν_{j_0} is small compared to ν (less than $\nu^{0.99}$, say), then this is not possible. To ensure that $\nu_{j_0} \gg \nu$, we want to truncate the sum in the LHS of (13) and keep only the terms with $\nu_{j_0} \asymp \nu$. This is why we have chosen k_ν with the property that it has rapid decay away from ν . As a bonus, we obtain an eigenfunction whose spectral parameter ν_{j_0} lies in an interval of bounded length around ν .

For the contributing hyperbolic classes, we will need additional arithmetic information:

Lemma 3.2. *Let $n \geq 1$ and $\eta \in R(n)$. Suppose that there exists $\gamma \in R^1$, $x \in \mathcal{F}$ and $\nu \geq 0$ with $x^{-1}\gamma^{-1}\eta\gamma x \in \text{supp } k_\nu$. Let $K = \mathbb{Q}(\eta) \subset B$ be the quadratic subfield generated by η , and let $\mathcal{O} = R \cap K$, which is an order in K . Then there exists a constant $C > 0$, independent of n and η such that $|\text{Tr } \bar{\rho}(\eta)| \leq C$ and the discriminant $D_{\mathcal{O}}$ of \mathcal{O} satisfies $|D_{\mathcal{O}}| \leq Cn \left| |\text{Tr } \bar{\rho}(\eta)| - 2 \right|$.*

Proof. Because $K \cap R \cong \gamma K \gamma^{-1} \cap R$, these orders have the same discriminant, and because $\text{Tr}(\gamma^{-1}\eta\gamma) = \text{Tr}(\eta)$ we may assume $\gamma = 1$. Then $\bar{\rho}(\eta)$ belongs to the bounded set $\mathcal{F}(\bigcup_\nu \text{supp } k_\nu) \mathcal{F}^{-1}$. Thus $\frac{1}{\sqrt{n}}\rho(\eta)$ belongs to a bounded set independent of n and ν . By the remarks in §2.1 there exists $C' > 0$ with $\|\frac{1}{\sqrt{n}}\eta\| \leq C'$, where $\|\cdot\|$ is the norm on $B \otimes_{\mathbb{Q}} \mathbb{R}$ we fixed in §2.2. Let $\alpha \in \mathcal{O}$ be such that $\alpha^2 = D_{\mathcal{O}}$. Then $\mathcal{O} \subset \frac{1}{2}\mathbb{Z} + \frac{1}{2}\alpha\mathbb{Z}$. Write $\eta = a + b\alpha$ with $a, b \in \frac{1}{2}\mathbb{Z}$. Then

$$|\text{Tr}(\eta)| = \|2a\| = \|\eta + \bar{\eta}\| \leq 2C' \sqrt{n}$$

so that $|\text{Tr}(\bar{\rho}(\eta))| \leq 2C'$. Because $\eta \notin \mathbb{Q}$ we have $b \neq 0$, so $|b| \geq 1/2$. From $n = N(\eta) = a^2 - D_{\mathcal{O}}b^2$ we have

$$|D_{\mathcal{O}}| = \frac{|a^2 - n|}{b^2} \leq 4|a^2 - n| = n(|\text{Tr } \bar{\rho}(\eta)| + 2) \left| |\text{Tr } \bar{\rho}(\eta)| - 2 \right|$$

and we can take $C = 2C' + 2$. \square

Let s denote the characteristic function of the squares, so that $|(R(n) \cap \mathbb{Q})/\pm 1| = s(n)$.

Proposition 3.3. *There exists an absolute constant $C > 0$ such that*

$$\sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) = s(n) \text{Vol}(\Gamma \backslash \mathfrak{h}) \cdot k_\nu(e) + O\left(n^{3/2} e^{C \log n / \log \log(n+1)}\right),$$

uniformly in $n \geq 1, \nu \geq 0$.

Proof. We may rewrite the sum in the RHS of (12) as

$$(14) \quad \frac{1}{2} \sum_{\mathcal{C}} \sum_{\eta \in \mathcal{C}} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx,$$

where the sum runs over all R^1 -conjugacy classes $\mathcal{C} \subset R(n) - \mathbb{Q}$. Fix \mathcal{C} and take $\eta \in \mathcal{C}$. Because B is a division algebra, $\bar{\rho}(\eta)$ is not parabolic.

When $\bar{\rho}(\eta)$ is hyperbolic, call $P > 1$ the square of its largest eigenvalue. Because η generates a real quadratic subfield $K \subset B$, the centralizer $Z_\Gamma(\bar{\rho}(\eta))$ is an infinite cyclic group. Let $\eta_0 \in K \cap R^1$ be totally positive and such that $\bar{\rho}(\eta_0)$ generates this centralizer, and let $P_0 > 1$ the square of the largest eigenvalue of $\bar{\rho}(\eta_0)$. By the computation in [18, §10.5] and our choice of k_ν ,

$$\begin{aligned} \sum_{\eta \in \mathcal{C}} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx &= \frac{\log P_0}{P^{1/2} - P^{-1/2}} \int_{\mathbb{R}} \widehat{k}_\nu(ir) e^{ir \log P} (2\pi)^{-1} dr \\ &\ll \frac{\log P_0}{P^{1/2} - 1}. \end{aligned}$$

We have

$$(P^{1/2} - 1)^2 \geq \frac{(P^{1/2} - 1)^2}{P^{1/2}} = P^{1/2} + P^{-1/2} - 2 = |\text{Tr}(\bar{\rho}(\eta))| - 2.$$

In order to bound P_0 , let $\sigma : K \rightarrow \mathbb{R}$ be an embedding and observe that $P_0^{1/2} + P_0^{-1/2} = \text{Tr}_{B/\mathbb{Q}}(\eta_0) = \sigma(\eta_0) + \sigma(\eta_0)^{-1}$. Thus $P_0 = \sigma(\eta_0)^{\pm 2}$, so that $\log P_0 = 2|\log \sigma(\eta_0)|$. When $R_{\mathcal{O}}$ is the regulator of the order $\mathcal{O} = R \cap K$, we have $|\log \sigma(\eta_0)| \in \{R_{\mathcal{O}}, 2R_{\mathcal{O}}\}$ according to whether the fundamental unit of \mathcal{O} has norm ± 1 . From the proof of Dirichlet's unit theorem, one has that $R_{\mathcal{O}}$ is bounded up to a constant by a power of the discriminant $D_{\mathcal{O}}$. Using the class number formula this can be improved to $D_{\mathcal{O}}^{1/2} \log D_{\mathcal{O}} \log \log D_{\mathcal{O}}$. (For non-maximal orders, this was done in [29].) We may now assume that the sum over $\eta \in \mathcal{C}$ is nonzero, in which case $D_{\mathcal{O}} \ll n(|\text{Tr} \bar{\rho}(\eta)| - 2) \ll n$ by Lemma 3.2. Combining this with the bound for $P^{1/2} - 1$, we obtain that for \mathcal{C} hyperbolic,

$$(15) \quad \sum_{\eta \in \mathcal{C}} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx \ll \frac{n^{1/2} (|\text{Tr} \bar{\rho}(\eta)| - 2)^{1/2} \log^2 n}{(|\text{Tr} \bar{\rho}(\eta)| - 2)^{1/2}} = n^{1/2} \log^2 n.$$

When $\bar{\rho}(\eta)$ is elliptic, let $0 < \theta \leq \pi/2$ be such that $\pm e^{i\theta}$ is an eigenvalue. The centralizer $Z_\Gamma(\bar{\rho}(\eta))$ is a finite cyclic group. By the computation in [18, §10.6],

$$\begin{aligned} \sum_{\eta \in \mathcal{C}} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx &= \frac{|Z_\Gamma(\bar{\rho}(\eta))|^{-1}}{\sin \theta} \int_{\mathbb{R}} \widehat{k}_\nu(ir) \frac{\cosh((\pi - 2\theta)r)}{\cosh(\pi r)} \frac{\pi}{2} dr \\ &\ll \frac{1}{\sin \theta}. \end{aligned}$$

Because $\bar{\rho}(\eta)$ is elliptic, $|\mathrm{Tr}(\eta)|^2 < 4n$, so $|\mathrm{Tr}(\eta)|^2 \leq 4n - 1$. We have

$$2 - |\mathrm{Tr}(\bar{\rho}(\eta))| \geq 2 - \sqrt{4 - \frac{1}{n}} = \frac{1}{n \left(2 + \sqrt{4 - \frac{1}{n}}\right)} \geq \frac{1}{4n}$$

so that

$$2 \sin^2 \theta = 2 - 2 \cos^2 \theta \geq 2 - 2 \cos \theta = 2 - |\mathrm{Tr}(\bar{\rho}(\eta))| \geq (4n)^{-1}.$$

We obtain that, for \mathcal{C} elliptic,

$$(16) \quad \sum_{\eta \in \mathcal{C}} \int_{\mathcal{F}} k_\nu(x^{-1}\eta x) dx \ll n^{1/2}.$$

We now count the number of contributing conjugacy classes. For the term in (14) corresponding to η to be nonzero, there must exist x in the compact set \mathcal{F} with $x^{-1}\bar{\rho}(\eta)x \in \mathrm{supp} k_\nu$. We may thus restrict the sum (14) to \mathcal{C} which contain an element η with $\|\eta\| \leq C'\sqrt{n}$. The number of such conjugacy classes is bounded by the cardinality of the set

$$M = \{\eta \in R(n) : \|\eta\| \leq C'\sqrt{n}\}.$$

This is a counting problem that has arisen many times in the context of sup norms of Maass cusp forms. Letting $(\eta_0, \eta_1, \eta_2, \eta_3)$ be a fixed \mathbb{Z} -basis of R and using that $\|\cdot\|$ is equivalent to the sup norm w.r.t. this basis, we see that $|M| \ll (\sqrt{n})^4$. A more careful treatment gives $|M| \ll n \exp(C \log n / \log \log(n+1))$ for some $C > 0$; see for example [22, §5] or Remark 4.4 below. Multiplying this bound for $|M|$ by the larger one of the bounds (15) and (16), the claim follows. \square

3.2. Relative pre-trace formula.

Proposition 3.4. *There exists a nonnegative $b \in C_c^\infty(\mathbb{R})$ depending only on Γ and L , such that for every $n \in \mathbb{N}_{>0}$,*

$$\begin{aligned} \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) \mathcal{P}(\phi_j)^2 &= |N_{R(n)}(F)/R_F^1| \cdot \mathrm{Vol}(\Gamma_L \backslash L) \cdot \int_{\mathbb{R}} k_\nu(a(t)) dt \\ &+ \sum_{\eta \in (R(n) - N_{R(n)}(F))/\pm 1} I(\nu, g_0^{-1}\eta g_0), \end{aligned}$$

where for $g \in G$ we define

$$(17) \quad I(\nu, g) = \int_{\mathbb{R} \times \mathbb{R}} b(s)b(t) k_\nu(a(-s)ga(t)) ds dt.$$

Proof. We start from the relative pre-trace formula (11). We have that L has unit length parametrization by $g_0 a(t) i$ where $a(t) = \exp\begin{pmatrix} t/2 & 0 \\ 0 & -t/2 \end{pmatrix}$. Let $\mathcal{F}_L \subset L$ be a fixed fundamental domain for the action of Γ_L . Write

$$\begin{aligned} \sum_{\eta \in R(n)/\pm 1} \int_{\mathcal{F}_L \times \mathcal{F}_L} k_\nu(x^{-1}\eta y) dx dy &= \sum_{\eta \in N_{R(n)}(F)/\pm 1} \int_{\mathcal{F}_L \times \mathcal{F}_L} k_\nu(x^{-1}\eta y) dx dy \\ &+ \sum_{\eta \in (R(n) - N_{R(n)}(F))/\pm 1} \int_{\mathcal{F}_L \times \mathcal{F}_L} k_\nu(x^{-1}\eta y) dx dy. \end{aligned}$$

The first term equals, by unfolding the sum in the second variable and then making the change of variables $y \leftarrow \eta^{-1}y$,

$$\begin{aligned}
 & \sum_{\eta \in N_{R(n)}(F)/R_F^1} \int_{\mathcal{F}_L \times L} k_\nu(x^{-1}\eta y) dx dy \\
 &= \sum_{\eta \in N_{R(n)}(F)/R_F^1} \int_{\mathcal{F}_L \times L} k_\nu(x^{-1}y) dx dy \\
 &= |N_{R(n)}(F)/R_F^1| \int_{[0, \text{Vol}(\mathcal{F}_L)] \times \mathbb{R}} k_\nu(a(-s+t)) ds dt \\
 &= |N_{R(n)}(F)/R_F^1| \cdot \text{Vol}(\Gamma_L \backslash L) \cdot \int_{\mathbb{R}} k_\nu(a(t)) dt
 \end{aligned}$$

For the second term, we unfold the sum as follows:

$$\begin{aligned}
 & \sum_{\delta \in \Gamma_L \backslash \overline{\rho}(R(n) - N_{R(n)}(F))/\Gamma_L} \sum_{\gamma \in \Gamma_L \delta \Gamma_L} \int_{\mathcal{F}_L \times \mathcal{F}_L} k_\nu(x^{-1}\gamma y) dx dy \\
 &= \sum_{\delta \in \Gamma_L \backslash \overline{\rho}(R(n) - N_{R(n)}(F))/\Gamma_L} \int_{L \times L} k_\nu(x^{-1}\delta y) dx dy
 \end{aligned}$$

Here, we used that for δ as above, the integral over $L \times L$ converges absolutely thanks to the compact support of k_ν , and that the map $\Gamma_L \times \Gamma_L \rightarrow G : (\gamma_1, \gamma_2) \mapsto \gamma_1 \delta \gamma_2$ is injective. Indeed, the contrary would imply that $\Gamma_L \cap \delta \Gamma_L \delta^{-1} \neq \{1\}$, so in particular $A_L \cap \delta A_L \delta^{-1} \neq \{1\}$. This would imply $A_L = \delta A_L \delta^{-1}$ so that $\delta \in N_G(A_L)$. If $\delta = \overline{\rho}(\eta)$ with $\eta \in R(n)$, this means that $\eta \in N_{R^+}(F)$, a contradiction.

Now we introduce a smooth cutoff function in each of the integrals. Let $b_0 \in C_0^\infty(L)$ be Γ_L -invariant and such that $\sum_{\gamma \in \Gamma_L} b_0(\gamma z) = 1$ for all $z \in L$. Then the sum equals

$$\begin{aligned}
 & \sum_{\delta \in \Gamma_L \backslash \overline{\rho}(R(n) - N_{R(n)}(F))/\Gamma_L} \int_{L \times L} \sum_{\gamma_1, \gamma_2 \in \Gamma_L} b_0(\gamma_1 x) b_0(\gamma_2 y) k_\nu(x^{-1}\delta y) dx dy \\
 & \sum_{\delta \in \Gamma_L \backslash \overline{\rho}(R(n) - N_{R(n)}(F))/\Gamma_L} \int_{L \times L} \sum_{\gamma_1, \gamma_2 \in \Gamma_L} b_0(x) b_0(y) k_\nu(x^{-1}\gamma_1^{-1} \delta \gamma_2 y) dx dy \\
 &= \sum_{\gamma \in (R(n) - N_{R(n)}(F))/\pm 1} \int_{L \times L} b_0(x) b_0(y) k_\nu(x^{-1}\gamma y) dx dy,
 \end{aligned}$$

where we made the change of variables $(x, y) \leftarrow (\gamma_1 x, \gamma_2 y)$ and merged the sum over δ with the sum over γ_1, γ_2 . We obtain the statement of the proposition, with $b(t) := b_0(g_0 a(t))$. \square

For the integral appearing in the main term in Proposition 3.4, we shall prove in §5:

Proposition 3.5. *We have*

$$\int_{\mathbb{R}} k_\nu(a(t)) dt \asymp 1$$

as $\nu \rightarrow \infty$.

Because the mass of k_ν is concentrated around K , we expect the integral $I(\nu, g)$ to be large when $a(-s)ga(t) \in K \backslash G / K$ is close to K on a set of large measure in $\text{supp } b \times \text{supp } b$. That is, when the geodesic segments $\{a(s)i : s \in \text{supp } b\}$ and $\{ga(t)i : t \in \text{supp } b\}$ are close to each other. We see that the size of $I(\nu, g)$ should be related to the distance of g to $N_G(A)$. This is quantified as follows:

Proposition 3.6. *Define $I(\nu, g)$ by (17).*

- (1) *There exists $C' > 0$ independent of $\nu \geq 0$ such that $I(\nu, g) = 0$ unless $d(g, e) \leq C'$.*
- (2) *For $g \in G$, we have*

$$|I(\nu, g)| \ll (1 + \nu \cdot d(g, N_G(A)))^{-1/2}$$

uniformly in g and $\nu \geq 0$.

Proof. We prove the first assertion and postpone the proof of the second assertion until §6. By construction, k_ν is supported on points at bounded distance from e . If $I(\nu, g)$ is nonzero, there exist $t, s \in \text{supp } b$ with $k_\nu(a(-t)ga(s)) \neq 0$. Then g belongs to the set $a(\text{supp } b)(\text{supp } k_\nu)a(-\text{supp } b)$, which is bounded uniformly in ν . \square

4. COUNTING HECKE RETURNS

4.1. Counting stabilizers. Let $F \subset B$ as before be the real quadratic number field corresponding to the geodesic L . We want to understand the factor $|N_{R(n)}(F)/R_F^1|$, which appears in the main term in the relative pre-trace formula. Denote by \mathfrak{f}_{R_F} the conductor of the order R_F , by P_{R_F} the set of principal ideals of the maximal order \mathcal{O}_F which are generated by an element of R_F , and by $P_{R_F}(n)$ the set of such ideals of norm n .

Lemma 4.1. *When $n \geq 1$ is coprime to \mathfrak{f}_{R_F} , we have*

$$|N_{R(n)}(F)/R_F^1| \geq |P_{R_F}(n)|$$

Proof. We may bound the LHS from below by $|R_F(n)/R_F^1|$. We show that for $(n, \mathfrak{f}_{R_F}) = 1$, this cardinality equals $|P_{R_F}(n)|$. An element $\eta \in R_F(n)$ determines a principal ideal $\eta R_F \subset R_F$, which determines a principal ideal $\eta \mathcal{O}_F \in P_{R_F}(n)$. The composition of the two maps obtained in this way, is by definition surjective onto $P_{R_F}(n)$. The first map has fibers which are full orbits under multiplication by R_F^1 . The second is injective, provided that n is coprime to the conductor \mathfrak{f}_{R_F} . \square

4.2. Bounding approximate stabilizers. In order to control the sum appearing in the error term in Proposition 3.4, we will need an upper bound for the number of elements $\eta \in R(n)$ that are close to stabilizing L without actually stabilizing it. That is, for the cardinality of the sets

$$M(n, \delta) = \{\eta \in R(n) : d(\bar{\rho}(\eta), e) \leq C', 0 < d(\bar{\rho}(\eta), N_G(A_L)) \leq \delta\},$$

where $C' > 0$ is an arbitrary constant which we fix throughout this section. This counting problem is similar to, but slightly different from the ones considered in [19, 21] in the context of upper bounds for sup norms resp. geodesic periods of Maass cusp forms: In the definition of $M(n, \delta)$, we are excluding the $\eta \in R(n)$ that stabilize L , so that the upper bound for $|M(n, \delta)|$ we obtain is smaller than what a direct invocation of [21, Lemma 3.3] would imply. This is necessary, because including those η in the definition of $M(n, \delta)$ would force any upper bound to be at least as large as $|N_{R(n)}(F)/R_F^1|$, while we want the latter quantity to dominate the

error term. Our method for bounding $|M(n, \delta)|$ however, uses many ideas which originate in [19].

Lemma 4.2. *There exists an absolute constant $C > 0$ such that when $D \in \mathbb{Z}_{>0}$ is a non-square, $0 < \delta \leq B$ and $n \geq 1$,*

$$\begin{aligned} & \#\{(u, v) \in \mathbb{Z}^2 : 0 < |u^2 - Dv^2 - n| \leq \delta n, |u|, |v| \leq Bn\} \\ & \ll \delta n e^{C \log n / \log \log(n+1)} \end{aligned}$$

where the implicit constant depends on D and B .

Proof. We want to estimate

$$\begin{aligned} & \#\{(u, v) \in \mathbb{Z}^2 : 0 < |u^2 - Dv^2 - n| \leq \delta n, |u|, |v| \leq Bn\} \\ & = \sum_{0 < |m-n| \leq \delta n} \#\{u, v : u^2 - Dv^2 = m, |u|, |v| \leq Bn\} \end{aligned}$$

Let $K = \mathbb{Q}(\sqrt{D})$. Fix m as in the sum. Any pair (u, v) as above determines an element $z = u + v\sqrt{D} \in K$ of norm m , and if $|\cdot|_1, |\cdot|_2$ denote the two Archimedean absolute values of K , we have $|z|_1, |z|_2 \leq Bn(1 + \sqrt{D})$. We obtain the upper bound

$$\begin{aligned} & \#\{(u, v) \in \mathbb{Z}^2 : u^2 - Dv^2 = m, |u|, |v| \leq Bn\} \\ & \leq \#\{z \in \mathcal{O}_K : N(z) = m, |z|_1, |z|_2 \leq Bn(1 + \sqrt{D})\}. \end{aligned}$$

The latter set maps to the set of integral ideals in \mathcal{O}_K with norm m . There are at most $\tau(m)$ such ideals. We may bound the divisor function $\tau(m)$ by its maximal order $\exp(C \log m / \log \log(m+1)) \ll_\delta \exp(C \log n / \log \log(n+1))$ for some $C > 0$. The fibers of said map are orbits under multiplication by units of norm 1. Let ε be the fundamental unit of K . The bound on $|z|_1$ and $|z|_2$ implies that the fibers of that map are of size at most

$$\ll \frac{\log(Bn(1 + \sqrt{D}))}{\log |\varepsilon|} \ll 1 + \log n$$

We obtain the upper bound

$$\begin{aligned} & \sum_{0 < |m-n| \leq \delta n} \#\{u, v : u^2 - Dv^2 = m, |u|, |v| \leq Bn\} \\ & \ll \delta n \cdot \exp(C \log n / \log \log(n+1)) \cdot (1 + \log n), \end{aligned}$$

as desired. Here we used that the number of terms in the sum is $\ll \delta n$, and not just $\ll \delta n + 1$, since we omit the term for $m = n$. \square

Lemma 4.3. *There exists an absolute constant $C > 0$ such that*

$$\#M(n, \delta) \ll \delta n e^{C \log n / \log \log(n+1)}.$$

Proof. Let $D > 0$ be the discriminant of F and take $\alpha \in F$ with $\alpha^2 = D$. For $t, u \in \mathbb{R}$ we have $N_{M_2(\mathbb{R})/\mathbb{R}}(t + u\rho(\alpha)) = t^2 - Du^2$, so that

$$(18) \quad A_L = \bar{\rho}((F \otimes_{\mathbb{Q}} \mathbb{R})^1) = \left\{ \overline{t + u\rho(\alpha)} : t^2 - Du^2 = 1 \right\}$$

Let $\omega \in R^+$ be the Skolem–Noether element from §2.3. Multiplying ω , if necessary, by an element of negative norm in F and by a suitable integer, we may assume that $\omega \in R^+$. Let $E := N_{B/\mathbb{Q}}(\omega) = -\omega^2 \in \mathbb{Z}_{>0}$. Let $S \subset B$ be the order $\mathbb{Z} + \mathbb{Z}\alpha + \mathbb{Z}\omega + \mathbb{Z}\alpha\omega$. Let $f \in \mathbb{N}_{>0}$ be such that $f \cdot R \subset S$. Now let $\eta \in M(n, \delta)$.

We can write $\eta = x_0 + x_1\alpha + x_2\omega + x_3\alpha\omega$ with $x_0, x_1, x_2, x_3 \in \frac{1}{f}\mathbb{Z}$, and we have $N_{B/\mathbb{Q}}(\eta) = x_0^2 - Dx_1^2 + Ex_2^2 - DEx_3^2 = n$.

Suppose first that $d(\bar{\rho}(\eta), A_L) \leq \delta$. Let C' be the constant from the beginning of the section, with respect to which $M(n, \delta)$ is defined. Because $M(n, \delta) = M(n, C')$ for $\delta \geq C'$, it is no restriction to assume $\delta \leq C'$. By the remarks in §2.1, from $d(\bar{\rho}(\eta), A_L) \leq \delta$ and $d(\bar{\rho}(\eta), e) \leq C'$ it follows that there exists $a \in \mathrm{SL}_2(\mathbb{R})$ with $\bar{a} \in A_L$ and $\|\frac{1}{\sqrt{n}}\rho(\eta) - a\| \ll \delta$. Here, the norm is any fixed norm on $M_2(\mathbb{R})$. Writing $a = t + u\rho(\alpha)$ as in (18), it follows that in $B \otimes_{\mathbb{Q}} \mathbb{R}$,

$$\frac{1}{\sqrt{n}}(\eta \otimes 1) = 1 \otimes t + \alpha \otimes u + O(\delta)$$

and comparing coordinates in the basis $(1, \alpha, \omega, \alpha\omega)$, we obtain

$$(19) \quad \begin{aligned} \frac{x_0}{\sqrt{n}} &= t + O(\delta) \\ \frac{x_1}{\sqrt{n}} &= u + O(\delta) \\ \frac{x_2}{\sqrt{n}} &= O(\delta) \\ \frac{x_3}{\sqrt{n}} &= O(\delta), \end{aligned}$$

so that

$$1 = t^2 - Du^2 = \frac{x_0^2}{n} - D\frac{x_1^2}{n} + O(\delta).$$

That is, $|x_0^2 - Dx_1^2 - n| \ll \delta n$. The assumption $d(\bar{\rho}(\eta), e) \leq C'$ implies that $x_0, x_1, x_2, x_3 \ll \sqrt{n}$. Because $d(\bar{\rho}(\eta), N_G(A)) > 0$, we have $\eta \notin F$, so that $(x_2, x_3) \neq (0, 0)$, so that $x_0^2 - Dx_1^2 = n - E(x_2^2 - Dx_3^2) \neq n$. We obtain that the integers (fx_0, fx_1) satisfy

$$0 < |(fx_0)^2 - D(fx_1)^2 - f^2n| \ll \delta n$$

and $|fx_0|, |fx_1| \ll n$. By Lemma 4.2 and the assumption that $\delta \leq C'$, the number of possible values for (x_0, x_1) is at most $\ll \delta n e^{C \log n / \log \log(n+1)}$ for some $C > 0$. If we fix x_0 and x_1 , then x_2 and x_3 satisfy

$$(20) \quad |(fx_2)^2 - D(fx_3)^2| = f^2/E \cdot |n - N_{B/\mathbb{Q}}(x_0 + x_1\alpha)| \ll n$$

and $|fx_2|, |fx_3| \ll n$. By counting ideals in $\mathbb{Q}(\sqrt{D})$ as in the proof of Lemma 4.2, we see that the number of (x_2, x_3) satisfying the equality in (20) can be bounded by $\ll \exp(C \log n / \log \log(n+1))$. This proves that the number of $\eta \in M(n, \delta)$ with $d(\bar{\rho}(\eta), A_L) \leq \delta$ is at most $\ll \delta n e^{2C \log n / \log \log(n+1)}$.

Suppose now that $\eta \in M(n, \delta)$ is such that $d(\bar{\rho}(\eta), N_G(A_L) - A_L) \leq \delta$. We have $\bar{\rho}(\omega) \in N_G(A_L) - A_L$, so we obtain that $d(\bar{\rho}(\omega\eta), A_L) \ll \delta$, $d(\bar{\rho}(\omega\eta), e) \ll 1$ and $\omega\eta \in R(n \cdot E)$. By the first case (with a different constant C' in the definition of $M(nE, \delta)$), the number of such η is bounded by $\ll \delta n E e^{C \log(nE) / \log \log(nE+1)}$ for some $C > 0$.

Adding the contributions from the two types of η , the claim follows. \square

Remark 4.4. (1) It is clear from the proofs of Lemma's 4.2 and 4.3 that when $M'(n, \delta)$ is the set obtained by removing the condition $0 < d(\bar{\rho}(\eta), N_G(A))$ in the definition of $M(n, \delta)$, we obtain the upper bound

$$\#M'(n, \delta) \ll (\delta n + 1)e^{C \log n / \log \log(n+1)}$$

for some $C > 0$. In particular,

$$\#\{\eta \in R(n) : d(\bar{\rho}(\eta), e) \leq C'\} = M'(n, C') \ll ne^{C \log n / \log \log(n+1)}.$$

- (2) Note that, as opposed to [21, Lemma 3.3], our bound for $M(n, \delta)$ is not uniform in the geodesic L , which explains why we are able to get a factor δ instead of only $\sqrt{\delta}$. But this improvement will not play a major role when we use this bound in the proof of Lemma 7.1 below. In fact, the larger part of our estimate of the error term in Proposition 3.4 will come from the η with $d(\bar{\rho}(\eta), N_G(A_L)) \asymp 1$, for which the higher power of δ in Lemma 4.3 gives no improvement.

5. STATIONARY PHASE

Proof of Proposition 3.5. Let $c : \mathbb{R} \rightarrow [0, 1]$ be smooth, compactly supported and such that $c(t) = 1$ when $a(t) \in \bigcup_{\nu \geq 0} \text{supp } k_\nu$. By (8) and Fubini,

$$\begin{aligned} \int_{\mathbb{R}} k_\nu(a(t)) dt &= \int_{\mathbb{R}} k_\nu(a(t)) c(t) dt \\ &= \int_{\mathbb{R}} \int_0^\infty \widehat{k}_\nu(ir) \varphi_{ir}(a(t)) \beta(r) c(t) dr dt \\ &= \int_0^\infty \widehat{k}_\nu(ir) \beta(r) \int_{\mathbb{R}} c(t) \varphi_{ir}(a(t)) dt dr \end{aligned}$$

because the compact support of c and the rapid decay of \widehat{k}_ν make the double integral absolutely convergent. Let

$$(21) \quad \mathcal{L}(r) = \int_{\mathbb{R}} c(t) \varphi_{ir}(a(t)) dt,$$

which is a real number. By Proposition 5.1 below, there exist $r_0, C > 0$ such that $\beta(r)\mathcal{L}(r) \in [C/2, 2C]$ for $r \geq r_0$. Write

$$\begin{aligned} &\int_0^\infty \widehat{k}_\nu(ir) \beta(r) \mathcal{L}(r) dr \\ &= \int_0^{r_0} \widehat{k}_\nu(ir) \beta(r) \mathcal{L}(r) dr + \int_{r_0}^\infty \widehat{k}_\nu(ir) \beta(r) \mathcal{L}(r) dr. \end{aligned}$$

For $\nu > r_0$, the first term is bounded by $r_0(\nu - r_0)^{-1} \cdot \sup_{r \in [0, r_0]} \beta(r) |\mathcal{L}(r)|$, and thus $o(1)$ as $\nu \rightarrow \infty$. Because $\beta(r)\mathcal{L}(r) \in [C/2, 2C]$ for $r \geq r_0$, the second term is positive and asymptotically equivalent to $\int_{r_0}^\infty \widehat{k}_\nu(ir) dr$. This integral is bounded from above by

$$\begin{aligned} &\ll \int_{r_0}^\infty (1 + |\nu - r|)^{-2} dr \\ &\leq \int_{-\infty}^\infty (1 + |r|)^{-2} dr \\ &\ll 1 \end{aligned}$$

and when $\nu > r_0$ it is bounded from below by

$$\int_\nu^{\nu+1} \widehat{k}_\nu(ir) dr \geq 1.$$

Hence $\int_{\mathbb{R}} k_\nu(a(t)) dt \asymp \int_{r_0}^\infty \widehat{k}_\nu(ir) dr \asymp 1$. \square

Proposition 5.1. *Define $\mathcal{L}(r)$ by (21). There exists $C > 0$ such that*

$$\mathcal{L}(r) = Cr^{-1} + O(r^{-2})$$

as $r \rightarrow \infty$.

Proof. We have [18, (1.43)]

$$\varphi_{ir} = \int_{S^1} (\cosh t + x_1 \sinh t)^{ir - \frac{1}{2}} d\mu(x)$$

where $S^1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ and the integral is w.r.t. the uniform measure. Hence

$$\mathcal{L}(r) = \int_{\mathbb{R}} \int_{S^1} c(t) (\cosh t + x_1 \sinh t)^{ir - \frac{1}{2}} d\mu(x) dt.$$

By the stationary phase theorem [34, §VIII.2], the asymptotic behavior of this integral is prescribed by the local properties of the integrand at the critical points of the phase function

$$\phi(x, t) = \log(\cosh t + x_1 \sinh t).$$

Lemma 5.2. *The critical points of ϕ are $((\pm 1, 0), 0)$, and they are nondegenerate. The Hessian in the (x_1, t) -coordinates at those points is of shape*

$$\mathbf{H} = \begin{pmatrix} 0 & 1 \\ 1 & * \end{pmatrix},$$

which has signature $(p, q) = (1, 1)$.

Proof. At the points where $x_0 = 0$ we have $\partial\phi/\partial t = \frac{\sinh t + x_1 \cosh t}{\cosh t + x_1 \sinh t} \neq 0$, because $|\cosh t| > |\sinh t|$. Hence they cannot be critical points. In the open set where $x_0 \neq 0$, we have $\partial\phi/\partial x_1 = \frac{\sinh t}{\cosh t + x_1 \sinh t}$. This is zero only when $t = 0$. We have $\partial\phi/\partial t = \frac{\sinh t + x_1 \cosh t}{\cosh t + x_1 \sinh t}$, which does not vanish when $t = 0$, unless $x_1 = 0$. Thus the only critical points are $((\pm 1, 0), 0)$.

We have $\phi(x, 0) = 0$, so that at the points where $t = 0$ one has $\partial^2\phi/\partial x_1^2 = 0$. At those points we have $\partial\phi/\partial t = x_1$, so that $\partial^2\phi/\partial t \partial x_1 = 1$. This proves that the Hessian at the critical points of ϕ is of shape \mathbf{H} . \square

Adding the contributions of the two critical points, stationary phase implies that $\mathcal{L}(r) = Cr^{-1} + O(r^{-2})$ with

$$C = 2 \cdot c(0) \cdot 2\pi \cdot |\det(\mathbf{H})|^{-1/2} \cdot e^{i\pi(p-q)/4} = 4\pi. \quad \square$$

6. ORBITAL INTEGRALS

We now prove Proposition 3.6. For the most part, this corresponds to the ‘frequency 0’ case in [21, §7], whose method we follow. Throughout, let C' be the constant from Proposition 3.6. Expanding k_ν in terms of spherical functions (see (8)), we have for $\nu \geq 0$ and $g \in G$,

$$I(\nu, g) = \int_0^\infty J(r, g) \widehat{k}_\nu(ir) \beta(r) dr$$

where

$$J(r, g) = \int_K \int_{\mathbb{R} \times \mathbb{R}} b(s) b(t) e^{(1/2 + ir)H(ka(-s)ga(t))} ds dt dk$$

and H is as in §2.5. We will use the stationary phase method to prove an upper bound for $J(r, g)$. Proposition 3.6 will then follow by integrating this bound against $\widehat{k}_\nu(ir)\beta(r)$.

For each $g \in G$, we have a map $\alpha_g : K \rightarrow K$ defined by requiring that

$$kg \in NA\alpha_g(k).$$

It is smooth in (g, k) because the Iwasawa decomposition $G \xrightarrow{\sim} N \times A \times K$ is smooth, and we have $\alpha_{gh} = \alpha_h \circ \alpha_g$. For $k \in K$ and $y, z \in G$ we have that (see e.g. [21, Lemma 6.2])

$$H(ky^{-1}z) = H(\alpha_{y^{-1}}(k)z) - H(\alpha_{y^{-1}}(k)y).$$

This allows us to separate variables inside the exponential in the definition of $J(r, g)$:

$$\begin{aligned} J(r, g) &= \int_K \int_{\mathbb{R} \times \mathbb{R}} b(s)b(t)e^{(1/2+ir)H(ka(-s)ga(t))} ds dt dk \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_{\mathbb{R} \times \mathbb{R}} b(s)b(t)e^{(1/2+ir)H(k(\sigma)a(-s)ga(t))} ds dt d\sigma \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_{\mathbb{R} \times \mathbb{R}} b'(s, \theta)b(s)b(t)e^{(1/2+ir)(H(k(\theta)ga(t)) - H(k(\theta)a(s)))} ds dt d\theta, \end{aligned}$$

by substituting $k(\theta) = \alpha_{a(-s)}(k(\sigma))$. Here, b' is a nonzero smooth function on $\mathbb{R} \times \mathbb{R}/2\pi\mathbb{Z}$, coming from the change of variables. This oscillating integral has phase function

$$(22) \quad \phi(s, t, \theta, g) := H(k(\theta)ga(t)) - H(k(\theta)a(s)).$$

Define $c(s, t, \theta, g) = (2\pi)^{-1}b'(s, \theta)b(s)b(t)\exp(\phi(s, t, \theta, g)/2)$ so that

$$J(r, g) = \int_0^{2\pi} \int_{\mathbb{R} \times \mathbb{R}} c(s, t, \theta, g)e^{ir\phi(s, t, \theta, g)} ds dt d\theta.$$

6.1. Critical points of ϕ . We analyze the critical points of ϕ for fixed θ and g . Let $\mathcal{P} = \mathbb{R}/2\pi\mathbb{Z} \times G$ and let $\mathcal{S} \subset \mathcal{P}$ be the closed subset consisting of ‘singular’ parameters (θ, g) for which one of the geodesics $k(\theta)Ai$ and $k(\theta)gAi$ is vertical. For every g , there are at most four values of θ for which $(\theta, g) \in \mathcal{S}$.

Proposition 6.1. *When $(\theta, g) \in \mathcal{S}$ is fixed, ϕ has no critical points. When $(\theta, g) \in \mathcal{P} - \mathcal{S}$, ϕ has a unique critical point with Hessian given in (s, t) -coordinates by*

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}.$$

Proof. When g and θ are fixed, we have that (s, t) is a critical point of ϕ iff

$$\left. \frac{\partial H(k(\theta)a(s))}{\partial s} \right|_{s=\xi_1(\theta)} = \left. \frac{\partial H(k(\theta)ga(t))}{\partial t} \right|_{t=\xi_2(\theta, g)} = 0.$$

That is, iff the geodesics $k(\theta)Ai$ and $k(\theta)gAi$ in \mathfrak{h} are not vertical, i.e., are half-circles, and their midpoints are $k(\theta)a(s)i$ resp. $k(\theta)ga(t)i$. (Recall that $H(g) = \log(\text{Im}(gi))$.) It is clear from the geometric interpretation that ϕ has no critical points when $(\theta, g) \in \mathcal{S}$, and has exactly one critical point when $(\theta, g) \in \mathcal{P} - \mathcal{S}$. Moreover, this critical point is then nondegenerate, because the above geodesics (which are half-circles) have nonzero (Euclidean) curvature at their midpoints. Finally, the shape of the Hessian is computed in the proof of [21, Lemma 7.10]. \square

When $(\theta, g) \in \mathcal{P} - \mathcal{S}$, let $(\xi_1(\theta), \xi_2(\theta, g))$ be the unique critical point of ϕ . The following lemma says that this point diverges to ∞ as (θ, g) approaches \mathcal{S} :

Lemma 6.2. *Let $R > 0$. Suppose $g \in G$ with $d(g, e) \leq C'$ and $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ are such that $|\xi_1(\theta)|, |\xi_2(\theta, g)| \leq R$. Then there exists $\delta = \delta(R, C') > 0$ such that $d((\theta, g), \mathcal{S}) > \delta$. In particular, the set*

$$\mathcal{P}_0 = \{(\theta, g) \in \mathcal{P} - \mathcal{S} : (\xi_1(\theta), \xi_2(\theta, g), \theta, g) \in \text{supp } c, d(g, e) \leq C'\}$$

is at positive distance from \mathcal{S} , and is compact.

Proof. For the sake of contradiction, suppose that $d((\theta, g), \mathcal{S})$ could be arbitrarily small. Then there exists a sequence (θ_n, g_n) for which $(\theta_n, g_n, \xi_1(\theta_n, g_n), \xi_2(\theta_n, g_n))$ converges in $\mathbb{R}/2\pi\mathbb{Z} \times G \times \mathbb{R}^2$, with a limit in $\mathcal{S} \times \mathbb{R}^2$. Call (θ, g, x, y) its limit. By continuity of $\partial H(k(\theta)a(s))/\partial s$ and $\partial H(k(\theta)ga(t))/\partial t$, we would have

$$\left. \frac{\partial H(k(\theta)a(s))}{\partial s} \right|_{s=x} = \left. \frac{\partial H(k(\theta)ga(t))}{\partial t} \right|_{t=y} = 0,$$

contradicting that one of the geodesics $k(\theta)Ai$, $k(\theta)gAi$ is vertical. It follows that $d(\mathcal{P}_0, \mathcal{S}) > 0$. Because \mathcal{P}_0 is bounded in \mathcal{P} and closed in $\mathcal{P} - \mathcal{S}$, it is compact. \square

For $(\theta, g) \in \mathcal{P} - \mathcal{S}$, define

$$(23) \quad \psi(\theta, g) = \phi(\xi_1(\theta), \xi_2(\theta, g), \theta, g).$$

Define also

$$c_1(\theta, g) = \begin{cases} \pi c(\xi_1(\theta), \xi_2(\theta, g), \theta, g) & : (\theta, g) \in \mathcal{P} - \mathcal{S}, \\ 0 & : (\theta, g) \in \mathcal{S}. \end{cases}$$

Lemma 6.2 implies that c_1 is smooth on \mathcal{P} . Let C' be as in Proposition 3.6 and \mathcal{P}_0 as in Lemma 6.2.

Lemma 6.3. *We have*

$$(24) \quad J(r, g) = r^{-1} \int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, g)} d\theta + O(r^{-2})$$

as $r \rightarrow \infty$, where the implicit constant remains bounded as long as $d(g, e) \leq C'$.

Proof. For fixed θ, g we apply the method of stationary phase in the variables s and t : Suppose $(\theta, g) \in \mathcal{P} - \mathcal{S}$. Then ϕ has a unique critical point with Hessian determinant $-1/4$. Stationary phase [34, §VIII.2] implies

$$(25) \quad \int_{\mathbb{R} \times \mathbb{R}} c(s, t, \theta, g) e^{ir\phi(s, t, \theta, g)} ds dt = r^{-1} e^{ir\psi(\theta, g)} c_1(\theta, g) + O(r^{-2}),$$

uniformly for (θ, g) in compact subsets of $\mathcal{P} - \mathcal{S}$. Suppose $(\theta, g) \in \mathcal{P} - \mathcal{P}_0$, so that ϕ has no critical point in the support of c . The absence of critical points implies

$$\int_{\mathbb{R} \times \mathbb{R}} c(s, t, \theta, g) e^{ir\phi(s, t, \theta, g)} ds dt \ll_N r^{-N},$$

uniformly for (θ, g) in compact subsets of $\mathcal{P} - \mathcal{P}_0$. Because $c_1(\theta, g) = 0$ in this case, we see that (25) still holds (after assigning any value to $\psi(\theta, g)$ when $(\theta, g) \in \mathcal{S}$). We may now take compact subsets of $\mathcal{P} - \mathcal{S}$ and of $\mathcal{P} - \mathcal{P}_0$ such that the union of the two contains $\mathbb{R}/2\pi\mathbb{Z} \times \{g \in G : d(g, e) \leq C'\}$. Integrating (25) w.r.t. θ then yields the desired estimate. \square

6.2. Critical points of ψ . In view of the expression (24) for $J(r, g)$, we analyze the critical points of ψ , for fixed g . When $(\theta, g) \notin \mathcal{S}$, we have [21, Proposition 7.2, Lemma 7.3] that θ is a critical point of ψ iff the geodesics $k(\theta)Ai$ and $k(\theta)gAi$, which by assumption are half-circles, are concentric. A critical point θ is degenerate iff these geodesics coincide, that is, iff $g \in N_G(A)$, in which case every θ for which $(\theta, g) \notin \mathcal{S}$ is a degenerate critical point.

Remark 6.4. For $g \notin N_G(A)$, the locations of the critical points of $\psi(-, g)$ can be described geometrically: as stated above, θ is a critical point iff the half-circles $k(\theta)Ai$ and $k(\theta)gAi$ are concentric, that is iff they share a common perpendicular vertical geodesic. In particular, the geodesics $k(\theta)Ai$ and $k(\theta)gAi$ must not intersect in $\mathfrak{h} \cup \mathbb{P}^1(\mathbb{R})$, that is, the geodesic gAi must have both endpoints in $\mathbb{R}_{>0}$ or in $\mathbb{R}_{<0}$. This condition on g is also sufficient for the existence of a critical point: Under this assumption on g , there are exactly two critical points which can be described as follows. Because Ai and gAi do not intersect in $\mathfrak{h} \cup \mathbb{P}^1(\mathbb{R})$, these geodesics lie at positive distance from each other, and this distance is realized in \mathfrak{h} by a pair of points on Ai and gAi . Let j be the geodesic carrying the geodesic segment joining those points. As a consequence of Gauss's lemma, j is then perpendicular to both geodesics. Because the sum of the angles of a hyperbolic quadrilateral is less than 360° , there can be no other geodesic that is perpendicular to both. Thus the critical points are the two values of θ for which $k(\theta)j$ is vertical.

To bound $J(r, g)$, we distinguish two cases, depending on whether g is close to $N_G(A)$ or at positive distance from it. For g away from $N_G(A)$, the absence of nondegenerate critical points of ψ implies:

Lemma 6.5. *When $d(g, N_G(A)) \geq \delta > 0$ and $d(g, e) \leq C'$, we have*

$$\int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, g)} d\theta \ll_\delta (1 + r \cdot d(g, N_G(A)))^{-1/2}$$

uniformly in g and $r \geq 0$.

Proof. Because ψ has no degenerate critical points when $g \notin N_G(A)$, we have $\max(|\partial\psi/\partial\theta|, |\partial^2\psi/\partial\theta^2|) \gg 1$ uniformly for g in compact subsets of $G - N_G(A)$ and for $\theta \in \text{supp } c_1(-, g)$. The Van der Corput lemma [34, §VIII.1, Proposition 2, Corollary] implies that

$$\int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, g)} d\theta \ll r^{-1/2}$$

as $r \rightarrow \infty$, uniformly for g in compact subsets of $G - N_G(A)$. \square

We may now restrict our attention to g that are close to $N_G(A)$. By (22) and (23),

$$\psi(\theta, g) := H(k(\theta)ga(\xi_2(\theta, g))) - H(k(\theta)a(\xi_1(\theta))).$$

From the characterization of $k(\theta)ga(\xi_2(\theta, g))i$ as the midpoint of the geodesic $k(\theta)gAi$, we see that ψ is right $N_G(A)$ -invariant in g . Thus in order to estimate the RHS of (24), we will first assume that g lies in a small neighborhood of the identity, and then translate the estimate to a small neighborhood of $N_G(A)$.

By explicating the order of vanishing of $\partial^2\psi(\theta, g)/\partial\theta^2$ as $d(g, A) \rightarrow 0$, an application of the Van der Corput lemma shows that:

Lemma 6.6. [21, Corollary 7.6] *There is an open neighborhood U of $e \in G$ such that for all $b \in C_c^\infty(\mathbb{R}/2\pi\mathbb{Z} - \{\theta : (\theta, g) \in \mathcal{S}\})$ and $g \in U$ we have*

$$\int_0^{2\pi} b(\theta) e^{ir\psi(\theta, g)} d\theta \ll (1 + r \cdot d(g, A))^{-1/2},$$

where the implicit constant remains bounded when $\text{supp}(b)$ stays at positive distance from the set $\{\theta : (\theta, g) \in \mathcal{S}\}$ and the derivatives of b up to a certain order remain bounded.

Corollary 6.7. *There exists an open neighborhood V of $N_G(A)$ in G such that when $g \in V$ and $d(g, e) \leq C'$,*

$$\int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, g)} d\theta \ll (1 + r \cdot d(g, N_G(A)))^{-1/2}$$

uniformly in g and $r \geq 0$.

Proof. Let U be the neighborhood from Lemma 6.6. Take $g_0 \in N_G(A)$. When $g \in Ug_0$, then Lemma 6.6 applied to gg_0^{-1} and $b(\theta) = c_1(\theta, g)$ shows that

$$\begin{aligned} \int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, g)} d\theta &= \int_0^{2\pi} c_1(\theta, g) e^{ir\psi(\theta, gg_0^{-1})} d\theta \\ &\ll_{g_0} (1 + r \cdot d(gg_0^{-1}, A))^{-1/2} \\ &\ll_{g_0} (1 + r \cdot d(g, N_G(A)))^{-1/2} \end{aligned}$$

uniformly in $g \in Ug_0$. We may now take a fixed subset $N_0 \subset N_G(A)$ such that the Ug_0 for $g_0 \in N_0$ form a locally finite cover of $N_G(A)$, and let $V = \bigcup_{g_0 \in N_0} Ug_0$. \square

Proof of Proposition 3.6. By the first part of Proposition 3.6, we may restrict to the $g \in G$ with $d(g, e) \leq C'$. The estimates from Lemma 6.5 and Corollary 6.7 may be plugged into Lemma 6.3 to obtain

$$|J(r, g)| \ll r^{-1} (1 + r \cdot d(g, N_G(A)))^{-1/2}$$

as $r \rightarrow \infty$, uniformly when $d(g, e) \leq C'$. On the one hand we have, by estimating $|J(r, g)|\beta(r) \ll 1$, that

$$\begin{aligned} I(\nu, g) &= \int_0^\infty J(r, g) \widehat{k}_\nu(ir) \beta(r) dr \\ &\ll \int_0^\infty \widehat{k}_\nu(ir) dr \\ &\ll 1. \end{aligned}$$

On the other hand, when $g \notin N_G(A)$,

$$\begin{aligned} I(\nu, g) &\ll \int_0^{\nu/2} \widehat{k}_\nu(ir) dr + \int_{\nu/2}^\infty (r \cdot d(g, N_G(A)))^{-1/2} \widehat{k}_\nu(ir) dr \\ &\ll_N \nu^{-N} + d(g, N_G(A))^{-1/2} \nu^{-1/2} \int_{\nu/2}^\infty \widehat{k}_\nu(ir) dr \\ &\ll d(g, N_G(A))^{-1/2} \nu^{-1/2}, \end{aligned}$$

where we have used that $1 \ll d(g, e)^{-1/2} \leq d(g, N_G(A))^{-1/2}$. Combining the two estimates, the claim follows. \square

7. PROOF OF THEOREM 1.1

7.1. Amplification. Let $(a(n))_{n \geq 1}$ be a sequence of nonnegative real numbers, supported on a finite set of integers $n \geq 1$ which are coprime to the discriminant Δ_R . Let $T = (\sum a_n T_n)^2$, and denote by $\widehat{T}(\phi_j) \geq 0$ the T -eigenvalue of ϕ_j . Define for $\nu \geq 0$,

$$Q(\nu, a) := \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) = B(a) \text{Vol}(\Gamma \backslash \mathfrak{h}) k_\nu(e) + O(R(a)) ;$$

$$Q_L(\nu, a) := \sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \mathcal{P}(\phi_j)^2 .$$

Our aim is to obtain good asymptotic estimates for these spectral sums.

Lemma 7.1. *Define the orbital integral $I(\nu, g)$ by (17). There exists an absolute constant $C > 0$, such that for $n \geq 1$ and $\nu \geq 0$ we have*

$$\sum_{\eta \in R(n) - N_{R^+}(F)} |I(\nu, g_0^{-1} \eta g_0)| \ll (1 + \nu)^{-1/2} n e^{C \log n / \log \log(n+1)} .$$

Proof. By Lemma 3.6, there exists $C' > 0$ independent of ν and n such that only the terms with $d(g_0^{-1} \eta g_0, e) \leq C'$ have a nonzero contribution. Cover the interval $[0, C']$ with the intervals $I_0 = [0, (1 + \nu)^{-1}]$, $I_k = [e^{k-1}(1 + \nu)^{-1}, e^k(1 + \nu)^{-1}]$ for $1 \leq k \leq \log(C'(1 + \nu)) + 1$. Define $M(n, \delta)$ as in §4.2, with respect to this value of C' . When $d(g_0^{-1} \eta g_0, N_G(A)) \in I_0$, we apply the bounds

$$I(\nu, g_0^{-1} \eta g_0) \ll 1 ,$$

$$\#M(n, (1 + \nu)^{-1}) \ll (1 + \nu)^{-1} n e^{C \log n / \log \log(n+1)}$$

from Proposition 3.6 and Lemma 4.3, which imply that the contribution of all η with $d(g_0^{-1} \eta g_0, N_G(A)) \in I_0$ is bounded by $\ll (1 + \nu)^{-1} n e^{C \log n / \log \log(n+1)}$ for some $C > 0$. When $d(g_0^{-1} \eta g_0, N_G(A)) \in I_k$, we apply the bounds

$$I(\nu, g_0^{-1} \eta g_0) \ll e^{-k/2} ,$$

$$\#M(n, e^k(1 + \nu)^{-1}) \ll e^k (1 + \nu)^{-1} n e^{C \log n / \log \log(n+1)} .$$

Summing over k , the contribution of the η with $d(g_0^{-1} \eta g_0, N_G(A)) \notin I_0$ is bounded by $\ll (1 + \nu)^{-1/2} n e^{C \log n / \log \log(n+1)}$. \square

Proposition 7.2. *There exists an absolute constant $C > 0$ such that for (a_n) as above and for all $\nu \geq 0$,*

$$Q(\nu, a) = B(a) \text{Vol}(\Gamma \backslash \mathfrak{h}) k_\nu(e) + O(R(a)) ,$$

$$Q_L(\nu, a) = B_L(a) \text{Vol}(\Gamma_L \backslash L) \int_{\mathbb{R}} k_\nu(a(t)) dt + O((1 + \nu)^{-1/2} R_L(a)) ,$$

where

$$\begin{aligned}
B(a) &:= \sum_{m,n} a_m a_n \sum_{d|m,n} d \cdot s(mn/d^2), \\
R(a) &:= \sum_{m,n} a_m a_n \sum_{d|m,n} d \cdot \left(\frac{mn}{d^2}\right)^{3/2} \cdot e^{C \log(mn/d^2) / \log \log(1+mn/d^2)}, \\
B_L(a) &:= \sum_{m,n} a_m a_n \sum_{d|m,n} d \cdot |N_{R(mn/d^2)}(F) / R_F^1|, \\
R_L(a) &:= \sum_{m,n} a_m a_n \sum_{d|m,n} d \cdot \frac{mn}{d^2} \cdot e^{C \log(mn/d^2) / \log \log(1+mn/d^2)}.
\end{aligned}$$

Proof. The first asymptotic formula follows from the recurrence relation (7) and Proposition 3.3. The second statement follows similarly from Proposition 3.4 and the estimate from Lemma 7.1. \square

In order to bound the tails of the spectral sums $Q(\nu, a)$ and $Q_L(\nu, a)$, we shall need:

Lemma 7.3. *Let (a_n) be as above. Then for $\nu \geq 0$,*

$$\begin{aligned}
\sum_{|\nu_j - \nu| \leq 1} \widehat{T}(\phi_j) &\ll (\nu + 1) \cdot B(a) + R(a), \\
\sum_{|\nu_j - \nu| \leq 1} \widehat{T}(\phi_j) \mathcal{P}(\phi_j)^2 &\ll B_L(a) + (\nu + 1)^{-1/2} R_L(a),
\end{aligned}$$

uniformly in (a_n) and ν .

Proof. For the first estimate, we start from Proposition 7.2 and plug in the bound $k_\nu(e) \asymp \nu$, to obtain

$$\sum_j \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \leq C_0 \cdot \nu \cdot B(a) + R(a),$$

for some $C_0 > 0$ and for $\nu \geq \nu_{(0)}$, where $\nu_{(0)} > 0$ depends on the choice of the family (k_ν) . Here, we used that $B(a), R(a) \geq 0$. The terms in the LHS are nonnegative. Because $\widehat{k}_\nu(i\nu_j) \geq 1$ when $|\nu_j - \nu| \leq 1$, the claim follows for $\nu \geq \nu_{(0)}$ by discarding the terms with $|\nu_j - \nu| > 1$.

To treat the case where $\nu \leq \nu_{(0)}$, one may either use Proposition 7.2 and observe that our specific construction of the family (k_ν) satisfies $k_\nu(e) \ll 1$ for $\nu \ll 1$, or use the following argument: for every $\nu \leq \nu_{(0)}$, we may find integers $n_1, n_2 \in [0, \nu_{(0)}]$ such that the set $\overline{B}(\nu, 1) \cap \{\nu_j : j \geq 0\}$ is contained in $\overline{B}(n_1, 1) \cup \overline{B}(n_2, 1)$. Proposition 7.2 applied to k_{n_1} and k_{n_2} gives

$$\begin{aligned}
\sum_{|\nu_j - \nu| \leq 1} \widehat{T}(\phi_j) &\leq \sum_j \widehat{k}_{n_1}(\nu_j) \widehat{T}(\phi_j) + \sum_j \widehat{k}_{n_2}(\nu_j) \widehat{T}(\phi_j) \\
&\ll \max_{n \in [0, \nu_{(0)}] \cap \mathbb{Z}} k_n(e) \cdot B(a) + R(a),
\end{aligned}$$

as desired. The second estimate in the statement is proven similarly, by using the bound $\int_{\mathbb{R}} k_\nu(a(t)) \ll 1$ from Proposition 3.5. \square

Proposition 7.4. *For (a_n) as above, $C \geq 1$ and $\nu > C$,*

$$\begin{aligned} \sum_{|\nu_j - \nu| \leq C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) - B(a) \operatorname{Vol}(\Gamma \backslash \mathfrak{h}) k_\nu(e) & \ll C^{-1} \nu B(a) + R(a), \\ \sum_{|\nu_j - \nu| \leq C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \mathcal{P}(\phi_j)^2 - B_L(a) \operatorname{Vol}(\Gamma_L \backslash L) \int_{\mathbb{R}} k_\nu(a(t)) dt & \ll C^{-1} B_L(a) + \nu^{-1/2} R_L(a), \end{aligned}$$

where the implicit constant is uniform in (a_n) , C and ν .

Proof. This will follow by combining Lemma 7.3 with the rapid decay of \widehat{k}_ν . By Proposition 7.2, for the first estimate it suffices to prove that

$$\sum_{|\nu_j - \nu| > C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \ll C^{-1} \nu B(a) + R(a).$$

When $\nu > C$, the condition $|\nu_j - \nu| > C$ is equivalent to $|\operatorname{Re}(\nu_j) - \nu| > C$. Indeed, when $\nu_j \in \mathbb{R}$ this is trivial, and when $\nu_j \in [-i/2, i/2]$, both conditions are satisfied. Consider first the sum over the set $\{j : \operatorname{Re}(\nu_j) < \nu - C\}$. We break it up into sums where $\operatorname{Re}(\nu_j)$ belongs to an interval of length 1: For $n \geq 0$, we have

$$\begin{aligned} \sum_{\nu - C - (n+1) \leq \operatorname{Re}(\nu_j) < \nu - C - n} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) & \ll \frac{1}{(C+n)^2} \sum_{|\nu_j - (\nu - C - n - \frac{1}{2})| \leq 1} \widehat{T}(\phi_j) \\ & \ll \frac{\nu B(a) + R(a)}{(C+n)^2}, \end{aligned}$$

where we use the rapid decay of \widehat{k}_ν and Lemma 7.3. Summing over integers $n \in [0, \nu - C]$, we find that the sum over $\operatorname{Re}(\nu_j) < \nu - C$ is bounded up to a constant by $C^{-1}(\nu B(a) + R(a))$. The sum over $\operatorname{Re}(\nu_j) > \nu + C$ is similarly bounded up to a constant by

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(\nu + C + n)B(a) + R(a)}{(C+n)^3} & \leq \sum_{n=0}^{\infty} \frac{\nu(C+n)B(a) + R(a)}{(C+n)^3} \\ & \ll C^{-1}(\nu B(a) + R(a)). \end{aligned}$$

Combining the two estimates, the first statement follows. For the second statement, it suffices to prove that

$$\sum_{|\nu_j - \nu| > C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \mathcal{P}(\phi_j)^2 \ll C^{-1} B_L(a) + \nu^{-1/2} R_L(a).$$

Using the rapid decay of \widehat{k}_ν and Lemma 7.3, we find that the sum over the j with $\operatorname{Re}(\nu_j) > \nu + C$ is bounded up to a constant by

$$\sum_{n=0}^{\infty} \frac{B_L(a) + \nu^{-1/2} R_L(a)}{(C+n)^2} \ll C^{-1} (B_L(a) + \nu^{-1/2} R_L(a)).$$

The sum over the j with $\operatorname{Re}(\nu_j) < \nu - C$ is bounded up to a constant by

$$\begin{aligned} & \sum_{n \leq \nu - C} \frac{B_L(a) + (|\nu - C - n| + 1)^{-1/2} R_L(a)}{(C + n)^2} \\ & \ll C^{-1} B_L(a) + R_L(a) \left(\sum_{C+n \leq \sqrt{\nu}} \frac{\nu^{-1/2}}{(C+n)^2} + \sum_{\sqrt{\nu} \leq C+n} \frac{1}{(C+n)^2} \right) \\ & \ll C^{-1} B_L(a) + \nu^{-1/2} R_L(a). \end{aligned}$$

The second statement follows. \square

7.2. Optimal resonators. Let $F \subset B$ as before be the field corresponding to the geodesic L , and let \mathfrak{f}_{R_F} be the conductor of R_F . We seek to apply Proposition 7.4 to a sequence (a_n) for which the quotient $B_L(a)/B(a)$ is large, and for which the sums $R(a)$ and $R_L(a)$ are small compared to $B(a)$ resp. $B_L(a)$. A similar optimization problem has been considered in [22] in the context of lower bounds for point evaluations of Maass cusp forms. Recycling some of the results there, we obtain in our situation:

Proposition 7.5 (Optimal resonators). *Let $M > 3$ be a real number. There exists a sequence (a_n) of nonnegative real numbers supported on integers $n \leq M$ coprime to Δ_R , such that the following hold:*

(1) *We have the lower bound*

$$(26) \quad \frac{B_L(a)}{B(a)} \geq \exp \left(2\sqrt{2} \sqrt{\frac{\log M}{\log \log M}} \left(1 + O \left(\frac{\log \log \log M}{\log \log M} \right) \right) \right).$$

(2) *There exists a constant $C'' > 0$ independent of M such that*

$$(27) \quad R(a) \ll M^3 e^{C'' \log M / \log \log M},$$

$$(28) \quad R_L(a) \ll M^2 e^{C'' \log M / \log \log M}.$$

(3) $B(a) \geq 1$.

Proof. We first reduce to the precise situation in [22]. When (a_n) is as in §7.1, define the sum $B_{R_F}(a)$ by replacing the set $N_{R(mn/d^2)}(F)/R_F^1$ by the set $P_{R_F}(mn/d^2)$ in the definition of $B_L(a)$. When (a_n) is supported on integers coprime to \mathfrak{f}_{R_F} , Lemma 4.1 implies $B_L(a) \geq B_{R_F}(a)$, so that $B_L(a)/B(a) \geq B_{R_F}(a)/B(a)$. The latter quantity is the one considered in [22].

When M is sufficiently large, we construct a sequence (a_n) as follows: Let $L = \sqrt{2 \log M \log \log M}$, and define a multiplicative function f by prescribing its values on prime powers by

$$f(p^n) := \begin{cases} \frac{L}{p^{\log p}} & \text{if } \chi_F(p) = 1, p \nmid \Delta_R \mathfrak{f}_{R_F}, n = 1, L^2 < p \leq \exp(\log^2 L), \\ 0 & \text{otherwise.} \end{cases}$$

Define $a_n = f(n)$ when $n \leq M$ and $a_n = 0$ otherwise. The proof of [22, Lemma 5] gives that $B_{R_F}(a)/B(a)$ is as large as the RHS in (26). (And it is shown that this sequence is optimal, in the sense that for every sequence (a_n) supported on integers coprime to $\Delta_R \mathfrak{f}_{R_F}$, the ratio $B_{R_F}(a)/B(a)$ is at most as large as the RHS in (26).)

We have $B(a) \geq a_1^2 = 1$. It remains to prove the bounds (27) and (28). Because (a_n) is supported on integers $n \leq M$,

$$\begin{aligned} R(a) &\ll M^3 e^{C \log(M^2)/\log \log(M^2)} \sum_{m,n} a_m a_n \sum_{d|m,n} \frac{1}{d^2} \\ &\ll M^3 e^{2C \log M/\log \log M} \left(\sum_n a_n \right)^2. \end{aligned}$$

Using Chebyshev's estimates, we have

$$\begin{aligned} \sum_n a_n &\leq \prod_p (1 + f(p)) \\ &\leq \exp \left(\sum_p f(p) \right) \\ &\leq \exp \left(L \cdot \sum_{L^2 < p} \frac{1}{p \log p} \right) \\ &\ll \exp(L/\log L) \\ &\ll \exp(\sqrt{\log M}) \end{aligned}$$

which proves (27). Similarly, the estimate

$$R_L(a) \ll M^2 e^{2C \log M/\log \log M} \sum_{m,n} a_m a_n \sum_{d \leq M} \frac{1}{d}$$

gives us (28). □

Armed with this resonator sequence, we are ready to prove Theorem 1.1.

Proof of Theorem 1.1. Let $\nu > 0$ be large. Let C'' be the constant from Proposition 7.5, choose any $A > C''/8$ and let $M = \nu^{1/4} e^{-A \log \nu / \log \log \nu}$. Let (a_n) be the corresponding sequence given by Proposition 7.5. We check that $R(a)$ and $R_L(a)$ are small compared to $B(a)$ and $B_L(a)$. From (27) we have $R(a) \ll_\epsilon \nu^{3/4+\epsilon}$, so that $R(a) = o(\nu B(a))$. From (28) we have

$$\begin{aligned} R_L(a) &\ll \nu^{1/2} e^{-2A \log \nu / \log \log \nu} e^{C'' \log M / \log \log M} \\ &\ll \nu^{1/2} e^{(-2A+C''/4)(\log \nu / \log \log \nu)(1+o(1))} \\ &\ll \nu^{1/2}, \end{aligned}$$

so that $\nu^{-1/2} R_L(a) = o(B_L(a))$. By Proposition 2.5 and Proposition 3.5, there exists $C \geq 1$ such that

$$\begin{aligned} C^{-1} \nu &\leq \frac{1}{2} \text{Vol}(\Gamma \backslash \mathfrak{h}) k_\nu(e), \\ C^{-1} &\leq \frac{1}{2} \text{Vol}(\Gamma_L \backslash L) \int_{\mathbb{R}} k_\nu(a(t)) dt, \end{aligned}$$

for ν sufficiently large. Proposition 7.4 applied to the sequence (a_n) and such C gives

$$\begin{aligned} \sum_{|\nu_j - \nu| \leq C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) &\asymp B(a)\nu, \\ \sum_{|\nu_j - \nu| \leq C} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \mathcal{P}(\phi_j)^2 &\asymp B_L(a). \end{aligned}$$

In particular, there must exist at least one $\nu_j \in [\nu - C, \nu + C]$ with $\mathcal{P}(\phi_j)^2 \gg B_L(a)/(\nu B(a))$. That is,

$$\begin{aligned} (\nu^{1/2} \mathcal{P}(\phi_j))^2 &\gg \frac{B_L(a)}{B(a)} \\ &\gg \exp\left(2\sqrt{2} \sqrt{\frac{\log M}{\log \log M}} \left(1 + O\left(\frac{\log \log \log M}{\log \log M}\right)\right)\right). \end{aligned}$$

We have $\log M = \frac{1}{4} \log \nu \cdot (1 + O(\log \log \nu / \log \nu))$, $\log \log M = \log \log \nu \cdot (1 + O(1/\log \nu))$ and $\log \log \log M \ll \log \log \log \nu$. Using this, we obtain for this particular j ,

$$(\nu^{1/2} \mathcal{P}(\phi_j))^2 \gg \exp\left(\sqrt{2} \sqrt{\frac{\log \nu}{\log \log \nu}} \left(1 + O\left(\frac{\log \log \log \nu}{\log \log \nu}\right)\right)\right).$$

Taking square roots and writing $\nu = \sqrt{\lambda - \frac{1}{4}}$, the theorem follows. \square

7.3. Extreme values of L -functions. We now explain how to deduce Corollary 1.2 from (the proof of) Theorem 1.1. There are two issues: In general, the RHS of the period formula (6) involves a sum of periods, rather than a single one. Second, Theorem 1.1 does not necessarily produce even ϕ_j .

Proof of Corollary 1.2. We use the notations from the statement. Let $B, R, \Gamma, \Lambda_{d_F}$ be as in Remark 2.3, and associate to f_j the Laplace–Hecke eigenfunction ϕ_j on $\Gamma \backslash \mathfrak{h}$. By our assumptions on N and d_F , B is a quaternion division algebra and R a maximal order. Let ϕ_0 be the constant function $\text{Vol}(\Gamma \backslash \mathfrak{h})^{-1/2}$. Then $(\phi_j)_{j \geq 0}$ is an orthonormal basis of $L^2(\Gamma \backslash \mathfrak{h})$ as in §2.6. In (6), take $f = f_j$. The product of the Gamma factors in the LHS of (6) is of size $\lambda_j^{-1/2}$ by Stirling’s formula, and we have the lower bound $L(1, \pi_j, \text{Ad}) \gg 1/\log(1 + \lambda_j)$ [14]. We have the factorization

$$L(s, \pi_j \times \pi_{\chi_0}) = L(s, \pi_j) L(s, \pi_j \times \omega_F).$$

Thus it suffices to give a lower bound for the RHS of (6), of the same quality as in Theorem 1.1.

Let $\sigma \in R$ be an element of norm -1 . In the amplified pre-trace formula (9), replace x by $\overline{\sigma x}$ (where the bar denotes complex conjugation) and add the resulting equality to (9) to obtain

$$2 \sum_{\phi_j \text{ even}} \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) \phi_j(x) \phi_j(y) = \sum_{\eta \in R(n)/\pm 1} (k_\nu(x^{-1}\eta y) + k_\nu((\overline{\sigma x})^{-1}\eta y))$$

For every pair (ℓ, ℓ') of closed geodesics in Λ_{d_F} , choose lifts L, L' and fundamental domains $\mathcal{F}_L, \mathcal{F}_{L'}$ and integrate over $\mathcal{F}_L \times \mathcal{F}_{L'}$ to obtain

$$\begin{aligned} 2 \sum_{\phi_j \text{ even}} \widehat{k}_\nu(i\nu_j) \widehat{T}_n(\phi_j) \mathcal{P}_\ell(\phi_j) \mathcal{P}_{\ell'}(\phi_j) &= \sum_{\eta \in R(n)/\pm 1} \int_{\mathcal{F}_L \times \mathcal{F}_{L'}} k_\nu(x^{-1}\eta y) dx dy \\ &+ \sum_{\eta \in R(n)/\pm 1} \int_{\mathcal{F}_L \times \mathcal{F}_{L'}} k_\nu((\overline{\sigma x})^{-1}\eta y) dx dy. \end{aligned}$$

Say $L = g_0 A i$ and $L' = g'_0 A i$, and call the first sum in the RHS $S(g_0, g'_0)$. The second sum is then $S(\sigma g_0 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, g'_0)$. We discuss the first term first; the second will be dealt with similarly. Let F, F' be the subfields of B associated to L, L' . In $S(g_0, g'_0)$, we isolate a main term, which is the sum over η with $\eta L' = L$, i.e. $\eta F' \eta^{-1} = F$. An unfolding argument as in the proof of Proposition 3.4 shows that the main term equals

$$\#(N_{F, F'}(n)/R_{F'}^1) \text{Vol}(\Gamma_L \backslash L) \int_{\mathbb{R}} k_\nu(a(t)) dt,$$

where $N_{F, F'} = \{\eta \in R^+ : \eta F' \eta^{-1} = F\}$. In the error term, we may again introduce nonnegative smooth cutoff functions and write it as

$$\sum_{\eta \in (R(n) - N_{F, F'})/\pm 1} I_{F, F'}(\nu, g_0^{-1} \eta g'_0),$$

where $I_{F, F'}(\nu, g_0)$ is defined similarly to (17), in terms of the chosen cutoff functions. This error term may be bounded as in Lemma 7.1. Here, the diophantine problem consists of counting $\eta \in R(n)$ that are at distance $\ll \sqrt{n}$ from 0 and are close to $N_{F, F'}$. Multiplication on the right by a fixed element $\eta_0 \in N_{F', F}$ gives an injection $N_{F, F'}(n) \rightarrow N_{R(n)N_{B/\mathbb{Q}}(\eta_0)}(F)$, reducing to the counting problem in §4.2. The analogue of Proposition 3.6 remains true; the only thing that changes in its proof is the amplitude function in the oscillating integral.

For the sum $S(\sigma g_0 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, g'_0)$ we do the same, except that L is now replaced by $\overline{\sigma L}$ and F is replaced by $\sigma F \sigma^{-1}$. Summing over all pairs (ℓ, ℓ') and taking an amplifier (a_n) as in §7.1, Proposition 7.4 becomes

$$\begin{aligned} (29) \quad \sum_{\substack{\phi_j \text{ even} \\ |\nu_j - \nu| \leq C}} \widehat{k}_\nu(i\nu_j) \widehat{T}(\phi_j) \left(\sum_{\ell \in \Lambda_{d_F}} \mathcal{P}_\ell(\phi_j) \right)^2 &= \text{Vol}(\Gamma_L \backslash L) B_{d_F}(a) \int_{\mathbb{R}} k_\nu(a(t)) \\ &+ O(C^{-1} B_{d_F}(a) + R_{d_F}(a)) \end{aligned}$$

where L is a lift of an arbitrary $\ell \in \Lambda_{d_F}$ (the volume is independent of the choice), $B_{d_F}(a)$ is defined similarly to $B_L(a)$ (see Proposition 7.2) and $B_{d_F}(a) \geq B_L(a)$. We conclude by choosing the amplifier as in Proposition 7.5, taking $C > 0$ large enough and comparing (29) to the standard pre-trace formula. (In the latter, we don't bother to extract the contribution of even ϕ_j , since we can discard the contribution of odd ones using positivity, in the endgame.) \square

Note that we have placed ourselves in a situation where the order R is maximal, in order to avoid having to worry about oldforms. When R is Eichler of arbitrary level N , it might be possible to extract the contribution of newforms, by writing

down the pre-trace formulas for Eichler orders R_d of levels $d \mid N$ and using Möbius-inversion. But the arithmetic of the cardinalities $\#(N_{F,F'}(n)/R_d^1)$ (which appear in the main term in (29)) as d varies, is intricate in general.

REFERENCES

- [1] M. Abert, N. Bergeron, and E. Le Masson. Eigenfunctions and random waves in the Benjamini–Schramm limit. preprint at <https://arxiv.org/abs/1810.05601>, 2018.
- [2] R. Aurich and F. Steiner. Exact theory for the quantum eigenstates of the Hadamard–Gutzwiller model. *Physica D*, 48(2–3):445–470, March 1991.
- [3] R. Aurich and F. Steiner. Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Physica D*, 64(1–3):185–214, April 1993.
- [4] V. G. Avakumović. Über die Eigenfunktionen auf geschlossenen Riemannschen Mannigfaltigkeiten. *Math. Zeitschr.*, 65:327–344, 1956.
- [5] M. V. Berry. Regular and irregular semiclassical wavefunctions. *J. Phys. A*, 10(12):2083–2091, 1977.
- [6] V. Blomer, É Fouvry, E. Kowalski, Ph. Michel, D. Milićević, and W. Sawin. *The second moment theory of families of L-functions*. Memoirs of the American Mathematical Society, to appear.
- [7] A. Bondarenko and K. Seip. Large greatest common divisor sums and extreme values of the Riemann zeta function. *Duke Math. J.*, 166(9):1685–1701, 2017.
- [8] F. Brumley and S. Marshall. Lower bounds for Maass forms on semisimple groups. *Compos. Math.*, to appear.
- [9] X. Chen and C. D. Sogge. On integrals of eigenfunctions over geodesics. *Proc. Am. Math. Soc.*, 143(1):151–161, 2015.
- [10] R. de la Bretèche and G. Tenenbaum. Sommes de Gál et applications. *Proc. London Math. Soc.*, 119(3):104–134, 2019.
- [11] S. Dyatlov and M. Zworski. Quantum ergodicity for restrictions to hypersurfaces. *Nonlinearity*, 26(1):35–52, 2013.
- [12] M. Eichler. *Lectures on modular correspondences*. Tata Institute of Fundamental Research, 1965.
- [13] D. W. Farmer, S. M. Gonek, and C. P. Hughes. The maximum size of L -functions. *J. Reine Angew. Math.*, 609:215–236, August 2007.
- [14] D. Goldfeld, J. Hoffstein, and D. Lieman. Appendix: An effective zero-free region. *Ann. Math.*, 140(1):177–181, July 1994.
- [15] D. A. Hejhal and B. N. Rackner. On the topography of Maass waveforms for $\mathrm{PSL}(2, \mathbb{Z})$. *Exp. Math.*, 1(4):275–305, 1992.
- [16] T. Hilberdink. An arithmetical mapping and applications to ω -results for the Riemann zeta function. *Acta Arith.*, 139(4):341–367, 2009.
- [17] L. Hörmander. The spectral function of an elliptic operator. *Acta Math.*, 121:193–218, 1968.
- [18] H. Iwaniec. *Spectral methods of automorphic forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, 2nd edition, 2002.
- [19] H. Iwaniec and P. Sarnak. L^∞ norms of eigenfunctions on arithmetic surfaces. *Ann. Math.*, 141:301–320, 1995.
- [20] J.-P. Kahane. *Some random series of functions*, volume 5 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2nd edition, 1985.
- [21] S. Marshall. Geodesic restrictions of arithmetic eigenfunctions. *Duke Math. J.*, 165(3), 2016.
- [22] D. Milićević. Large values of eigenfunctions on arithmetic hyperbolic surfaces. *Duke Math. J.*, 155(2):365–401, 2010.
- [23] D. Milićević. Large values of eigenfunctions on arithmetic hyperbolic 3-manifolds. *Geom. Funct. Anal.*, 21(6):1375–1418, December 2011.
- [24] A. A. Popa. Central values of Rankin L -series over real quadratic fields. *Compos. Math.*, 142(4):811–866, July 2006.
- [25] M. Ratner. The central limit theorem for geodesic flows on n -dimensional manifolds of negative curvature. *Isr. J. Math.*, 16:181–197, 1973.
- [26] A. Reznikov. A uniform bound for geodesic periods of eigenfunctions on hyperbolic surfaces. *Forum Math.*, 27(3):1569–1590, 2015.

- [27] Z. Rudnick and P. Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Commun. Math. Phys.*, 161:195–213, 1994.
- [28] R. Salem and A. Zygmund. Some properties of trigonometric series whose terms have random signs. *Acta Math.*, 91:245–301, 1954.
- [29] J. W. Sands. Generalization of a theorem of Siegel. *Acta Arith.*, 58(1):47–56, 1991.
- [30] P. Sarnak. Arithmetic quantum chaos. In *The Schur lectures*, volume 8 of *Israel mathematical conference proceedings*, 1995.
- [31] P. Sarnak. Reciprocal geodesics. *Clay Math. Proc.*, 7:217–237, 2007.
- [32] Ya. G. Sinai. The central limit theorem for geodesic flows on manifolds of constant negative curvature. *Proc. USSR Acad. Sci.*, 133(6):1303–1306, 1960. in Russian.
- [33] K. Soundararajan. Extreme values of zeta and L -functions. *Math. Ann.*, 342:467–486, 2008.
- [34] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton University Press, 1993.
- [35] J. A. Toth and S. Zelditch. Quantum ergodic restriction theorems. I: interior hypersurfaces in domains with ergodic billiards. *Ann. Henri Poincaré*, 13(4):599–670, May 2012.
- [36] J. A. Toth and S. Zelditch. Quantum ergodic restriction theorems: manifolds without boundary. *Geom. Funct. Anal.*, 23(2):715–775, April 2013.
- [37] M. Vignéras. *Arithmétique des algèbres de quaternions*. Number 800 in *Lecture Notes in Mathematics*. Springer-Verlag, 1980.
- [38] J.-L. Walspurger. Sur les valeurs de certaines fonctions l automorphes et leur centre de symétrie. *Compos. Math.*, 54:173–242, 1985.
- [39] M. Young. The quantum unique ergodicity conjecture for thin sets. *Adv. Math.*, 286:958–1016, January 2016.
- [40] M. Young. Equidistribution of Eisenstein series on geodesic segments. *Adv. Math.*, 340:1166–1218, 2018.
- [41] S. Zelditch. Kuznecov sum formulae and Szegő limit formulae on manifolds. *Comm. Part. Differ. Equat.*, 17(1–2):221–260, 1992.
- [42] S. Zelditch. Quantum ergodicity and mixing of eigenfunctions. In *Encyclopedia of mathematical physics*, pages 183–196, 2006.
- [43] S. Zhang. Gross–Zagier formula for GL_2 . *Asian J. Math.*, 5(2):183–290, 2001.