Math. Ann. (2018) 371:1497–1543
https://doi.org/10.1007/s00208-018-1677-9

Mathematische Annalen

Equidistribution in shrinking sets and L^4-norm bounds for automorphic forms

Peter Humphries[1]

Received: 9 November 2017 / Revised: 6 April 2018 / Published online: 24 April 2018

Abstract We study two closely related problems stemming from the random wave conjecture for Maaß forms. The first problem is bounding the L^4-norm of a Maaß form in the large eigenvalue limit; we complete the work of Spinu to show that the L^4-norm of an Eisenstein series $E(z, 1/2+it_g)$ restricted to compact sets is bounded by $\sqrt{\log t_g}$. The second problem is quantum unique ergodicity in shrinking sets; we show that by averaging over the centre of hyperbolic balls in $\Gamma\backslash\mathbb{H}$, quantum unique ergodicity holds for almost every shrinking ball whose radius is larger than the Planck scale. This result is conditional on the generalised Lindelöf hypothesis for Hecke–Maaß eigenforms but is unconditional for Eisenstein series. We also show that equidistribution for Hecke–Maaß eigenforms need not hold at or below the Planck scale. Finally, we prove similar equidistribution results in shrinking sets for Heegner points and closed geodesics associated to ideal classes of quadratic fields.

Mathematics Subject Classification 11F12 (primary); 58J51 (secondary)

1 Introduction

1.1 Randomness of Maaß newforms

1.1.1 Random wave conjecture

Let $\mathcal{B}_0(\Gamma)$ denote the set of Hecke–Maaß eigenforms of weight zero and level 1 on the modular surface $\Gamma\backslash\mathbb{H}$, where $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ and $\mathbb{H}$ denotes the upper half-plane; we normalise $g \in \mathcal{B}_0(\Gamma)$ to be such that

Communicated by Kannan Soundararajan.

✉ Peter Humphries
pclhumphries@gmail.com

[1] Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK

$$\langle g, g\rangle := \int_{\Gamma\backslash\mathbb{H}} |g(z)|^2 \, d\mu(z) = 1,$$

where $d\mu(z) = y^{-2}\, dx\, dy$. A well-known conjecture of Berry [1] and Hejhal and Rackner [20] states that a Hecke–Maaß eigenform $g \in \mathcal{B}_0(\Gamma)$ of large Laplacian eigenvalue $\lambda_g = 1/4 + t_g^2$ ought to behave like a random wave. Here by a random wave, we mean a function of the form

$$g_\lambda(z) = \sum_{\lambda \leq \lambda_f \leq \lambda + \eta(\lambda)} c_f f(z),$$

where $\eta(\lambda) \to \infty$ as $\lambda \to \infty$ and $\eta(\lambda) = o(\lambda)$, each f is a normalised Hecke–Maaß eigenform, and the coefficients c_f are independent Gaussian random variables of mean 0 and variance 1. These are a randomised model of eigenfunctions of the Laplacian in the large eigenvalue limit $\lambda \to \infty$, and it is easier to prove (almost surely) results for random waves than for true eigenfunctions.

For $\Gamma\backslash\mathbb{H}$, there are situations in which random waves do not behave precisely like Laplacian eigenfunctions: random waves satisfy $\sup_{z\in K} |g_\lambda(z)| \asymp_K \sqrt{\log\lambda}$ almost surely for every compact subset K, whereas Milićević [41, Theorem 1] proved the existence of a dense subset of points $z \in \Gamma\backslash\mathbb{H}$ for which a subsequence of Hecke–Maaß eigenforms $g \in \mathcal{B}_0(\Gamma)$ may be much larger. Nonetheless, it is conjectured that Laplacian eigenfunctions should, on the whole, be well-modelled by random waves. This (admittedly loosely defined) conjecture is known as the random wave conjecture.

In this paper, we study two aspects of this conjecture: bounds for the L^4-norm of an automorphic form, and quantum unique ergodicity in shrinking balls. The former is a special case of the Gaussian moments conjecture, while the latter is a refinement of quantum unique ergodicity.

1.1.2 Gaussian moments conjecture

A particular manifestation of the random wave conjecture states that the moments of a Hecke–Maaß eigenform $g \in \mathcal{B}_0(\Gamma)$ should be identical to those of a Gaussian random variable in the large eigenvalue limit.

Conjecture 1.1 (Gaussian Moments Conjecture) *Let K be any fixed compact continuity set of $\Gamma\backslash\mathbb{H}$, so that the boundary of K has μ-measure zero, and let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. Then for every nonnegative integer n,*

$$\frac{1}{\mathrm{Var}_K(g)^{n/2}\,\mathrm{vol}(K)} \int_K g(z)^n \, d\mu(z) \tag{1.2}$$

converges to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}}\, dx = \begin{cases} \dfrac{2^{n/2}}{\sqrt{\pi}}\Gamma\left(\dfrac{n+1}{2}\right) & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd,} \end{cases}$$

as t_g tends to infinity. Here

$$\mathrm{Var}_K(g) := \frac{1}{\mathrm{vol}(K)} \int_K |g(z)|^2 \, d\mu(z).$$

When K is replaced by a noncompact set, the Gaussian moments conjecture ought not necessarily to hold for high moments. As explained in [21, Sect. 4], using a heuristic appearing in [19, Sect. 7], the transition range of the Whittaker function leads to a "tidal pulse" phenomenon near the cusp of $\Gamma\backslash\mathbb{H}$; when K is replaced by $\Gamma\backslash\mathbb{H}$, so that $\mathrm{Var}_{\Gamma\backslash\mathbb{H}}(g) = \mathrm{vol}\,(\Gamma\backslash\mathbb{H})^{-1}$, one can thereby show that there exists a subsequence of Hecke–Maaß eigenforms $g \in \mathcal{B}_0(\Gamma)$ for which (1.2) grows like a power of t_g whenever $n \geq 12$ is even. This is closely related to the fact that there exists a subsequence of Hecke–Maaß eigenforms for which

$$\|g\|_\infty \gg_\varepsilon t_g^{\frac{1}{6}-\varepsilon}.$$

Nonetheless, it is not unreasonable to conjecture that the Gaussian moments conjecture holds for smaller moments when K is replaced by $\Gamma\backslash\mathbb{H}$. Indeed, the conjecture holds by definition for $n \in \{0, 2\}$ and is easily shown to also be true when $n = 1$, as both sides vanish, while for $n = 3$, this can be shown to hold via the work of Watson [46].

1.1.3 Quantum unique ergodicity

Another manifestation of the randomness of Hecke–Maaß eigenforms is quantum unique ergodicity.

Conjecture 1.3 (quantum unique ergodicity in configuration space) *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. Then the probability measure $|g(z)|^2 \, d\mu(z)$ converges in distribution to the uniform probability measure on $\Gamma\backslash\mathbb{H}$ as t_g tends to infinity, so that for every continuity set $B \subset \Gamma\backslash\mathbb{H}$,*

$$\int_B |g(z)|^2 \, d\mu(z) = \frac{\mathrm{vol}(B)}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} + o_B(1)$$

as t_g tends to infinity.

By the Portmanteau theorem, this conjecture is equivalent to

$$\int_{\Gamma\backslash\mathbb{H}} f(z)|g(z)|^2 \, d\mu(z) = \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} \int_{\Gamma\backslash\mathbb{H}} f(z) \, d\mu(z) + o_f(1) \tag{1.4}$$

for every bounded continuous function on $\Gamma\backslash\mathbb{H}$.

It behoves us to mention that there is a stronger formulation of quantum unique ergodicity, namely quantum unique ergodicity in phase space, which is the cosphere bundle $S^*\,(\Gamma\backslash\mathbb{H}) \cong \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$: not only should the sequence of probability measures

$|g(z)|^2\, d\mu(z)$ equidistribute on the configuration space $\Gamma\backslash\mathbb{H}$, but that a microlocal lift of these measures to Wigner distributions on phase space should equidistribute with respect to the Liouville measure.

Quantum unique ergodicity in phase space, and hence also in configuration space, is known to be true via the work of Lindenstrauss [36] and Soundararajan [44]. However, this proof does not quantify the rate of equidistribution; in particular, it does not give explicit rates of decay for the terms

$$\int_{\Gamma\backslash\mathbb{H}} f(z)|g(z)|^2\, d\mu(z) \tag{1.5}$$

for fixed $f \in C_b(\Gamma\backslash\mathbb{H})$ as t_g tends to infinity. Watson [46, Corollary 1] has shown that optimal decay rates for these integrals follow directly from the generalised Lindelöf hypothesis.

The $n = 2$ case of the Gaussian moments conjecture for the set $K = \Gamma\backslash\mathbb{H}$—namely the L^4-norm of g—shares many similarities with quantum unique ergodicity in configuration space. In fact, it is extremely closely related to a more refined version of quantum unique ergodicity, namely equidistribution on shrinking sets.

1.1.4 Randomness of Eisenstein series

The Gaussian moments conjecture and quantum unique ergodicity ought to be true, once suitably modified, when $g(z) = E(z, 1/2 + it_g)$ is an Eisenstein series. Eisenstein series are not square-integrable, so one must use some sort of regularisation. One method is to use Zagier's regularisation of divergent integrals [50]; another is to replace $E(z, 1/2 + it_g)$ with the truncated Eisenstein series $\Lambda^T E(z, 1/2 + it_g)$ for some $T \geq 1$; this is defined for $\Re(s) > 1$ by

$$\Lambda^T E(z,s) := E(z,s) - \sum_{\substack{\gamma\in\Gamma_\infty\backslash\Gamma \\ \Im(\gamma z)>T}} \left(\Im(\gamma z)^s + \frac{\Lambda(2-2s)}{\Lambda(2s)}\Im(\gamma z)^{1-s}\right)$$

and extended by meromorphic continuation to the complex plane; here $\Lambda(s)$ denotes the completed Riemann zeta function.

For quantum unique ergodicity, we need not deal with the truncated version of the Eisenstein series provided that we take into account the growth of the L^2-norm of an Eisenstein series on compact sets.

Theorem 1.6 (Luo–Sarnak [39, Theorem 1.1]) *For any compact continuity set $K \subset \Gamma\backslash\mathbb{H}$ and for $g(z) = E\left(z, 1/2 + it_g\right)$,*

$$\int_K |g(z)|^2\, d\mu(z) = \frac{\log\left(\frac{1}{4} + t_g^2\right)\operatorname{vol}(K)}{\operatorname{vol}\left(\Gamma\backslash\mathbb{H}\right)} + o_K\left(\log t_g\right)$$

as t_g tends to infinity.

Since K is compact, one can replace $g(z)$ with $\Lambda^T E\left(z, 1/2 + it_g\right)$ for some T sufficiently large dependent on K. The presence of $\log(1/4 + t_g^2)$ essentially stems from the Maaß–Selberg relation; see Corollary 2.3.

Quantum unique ergodicity in phase space is also known for Eisenstein series; this is a result of Jakobson [30, Theorem 1].

1.2 The L^4-norm problem

The L^4-norm problem for a Hecke–Maaß eigenform g is the second nontrivial case of the Gaussian moments conjecture.

Conjecture 1.7 (L^4-norm problem) *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. As t_g tends to infinity,*

$$\int_{\Gamma\backslash\mathbb{H}} |g(z)|^4 \, d\mu(z) = \frac{3}{\operatorname{vol}(\Gamma\backslash\mathbb{H})} + o(1).$$

A similar statement can be formulated when g is an Eisenstein series, though some care must be taken, since Eisenstein series are not square-integrable; see [9].

In general, an unconditional proof of the L^4-norm problem seems quite difficult. A weaker conjecture (see, for example, [43, Conjecture 4]) is that

$$\|g\|_{L^4(\Gamma\backslash\mathbb{H})}^4 \ll_\varepsilon t_g^\varepsilon. \tag{1.8}$$

In certain special cases, this has been shown: when g is a dihedral Maaß eigenform, this is a result of Luo [38], while when g is a truncated Eisenstein series, this is a result of Spinu [45] (with the implicit constant of course dependent on the truncation parameter T).

Buttcane and Khan [6, Theorem 1.1] have recently given a proof, conditional on the generalised Lindelöf hypothesis, of the L^4-norm problem for a Hecke–Maaß eigenform $g \in \mathcal{B}_0(\Gamma)$. Our first main result is to give an unconditional upper bound for the L^4-norm of a truncated Eisenstein series that is sharper than (1.8).

Theorem 1.9 *Let $g(z) = \Lambda^T E\left(z, 1/2 + it_g\right)$. We have that*

$$\|g\|_{L^4(\Gamma\backslash\mathbb{H})}^4 \ll_T \left(\log t_g\right)^2.$$

Up to the implicit constant, Theorem 1.9 should be sharp, for the Maaß–Selberg relation implies that

$$\|g\|_{L^2(\Gamma\backslash\mathbb{H})}^4 = \left(\log\left(\left(\frac{1}{4} + t_g^2\right) T^2\right) + O\left(\left(\log t_g\right)^{2/3} \left(\log\log t_g\right)^{1/3}\right)\right)^2.$$

Remark 1.10 Theorem 1.9 was previously claimed by Spinu [45, Theorem 1.2], as was a proof of (1.8) for Hecke–Maaß cusp forms by Sarnak and Watson [43, Theorem 3]; in both cases, however, the proofs are incomplete, as we shall discuss further in Remark 3.3.

Remark 1.11 Djanković and Khan [9] have recently reformulated the L^4-norm problem for Eisenstein series by studying a regularised fourth moment of an Eisenstein series in the sense of Zagier [50]; cf. Sect. 2.2. This has the advantage that one ought to be able to prove an asymptotic for this regularised fourth moment, whereas Theorem 1.9 only provides an upper bound for the fourth moment of a truncated Eisenstein series.

1.3 Quantum unique ergodicity in shrinking sets

A natural strengthening of quantum unique ergodicity is to determine whether equidistribution still occurs if we vary the set B with t_g; in particular, if the size of B shrinks as t_g increases. This small scale equidistribution should be thought of as a reinterpretation of determining the rate of equidistribution, as opposed to determining explicit rates of decay for the terms in (1.5). Proving equidistribution in shrinking sets has applications towards bounds for the L^p-norms and size of nodal domains of eigenfunctions of the Laplacian; see [22].

We denote by $B = B_R(w)$ the hyperbolic ball of radius R centred at $w \in \Gamma\backslash\mathbb{H}$: its hyperbolic volume is

$$\mathrm{vol}\,(B_R) = 4\pi \sinh^2 \frac{R}{2},$$

which is independent of the centre w.

Question 1.12 *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. For what conditions on R, with regards to t_g, is it still true that*

$$\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) = \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} + o_w(1) \tag{1.13}$$

as t_g tends to infinity?

In the general setting of negatively curved manifolds, this question has independently been answered by Han [16, Theorem 1.5] and Hezari and Rivière [22, Proposition 2.1] for a full density subsequence of Laplacian eigenfunctions with the radius R shrinking at a rate $(\log \lambda_g)^{-\beta}$ for a particular range of $\beta > 0$ dependent on the manifold.

We should not expect equidistribution to hold when $R \ll t_g^{-1}$; indeed, Hejhal and Rackner [20, Sect. 5], writing Ψ_n in place of g, λ_n in place of $\lambda_g = 1/4 + t_g^2$, and A in place of R, state that

> …in the physics literature, $c/\sqrt{\lambda_n}$ is commonly referred to as the de Broglie wavelength. At length scales below $c/\sqrt{\lambda_n}$, one expects the topography of Ψ_n to look "essentially sinusoidal", that is, regular. It is only when A is substantially

bigger than the de Broglie wavelength that one stands any chance of seeing any type of Gaussian distribution.

We confirm this statement by showing that if $R \ll_A t_g^{-1}(\log t_g)^A$ for any $A > 0$, then there exist infinitely many points $w \in \Gamma\backslash\mathbb{H}$ for which (1.13) does *not* hold, so that the sequence of probability measures $|g(z)|^2\,d\mu(z)$ does not equidistribute on the shrinking balls of radius $t_g^{-1}(\log t_g)^A$ centred at these points. We think of $R \asymp t_g^{-1}$ as being the Planck scale, so that equidistribution need not occur within a logarithmic window of the Planck scale.

Theorem 1.14 *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. For every fixed Heegner point $w \in \Gamma\backslash\mathbb{H}$, we have that*

$$\frac{1}{\mathrm{vol}\,(B_R)}\int_{B_R(w)} |g(z)|^2\,d\mu(z) = \Omega\left(\exp\left(2\sqrt{\frac{\log t_g}{\log\log t_g}}\left(1 + O\left(\frac{\log\log\log t_g}{\log\log t_g}\right)\right)\right)\right)$$

for $R \ll_A t_g^{-1}(\log t_g)^A$ for any $A > 0$ as t_g tends to infinity.

Nevertheless, we should expect equidistribution to occur at every scale larger than the Planck scale, namely $R \gg t_g^{-\delta}$ for any $\delta < 1$. Towards this, Young [47] has proved the following.

Theorem 1.15 (Young [47, Proposition 1.5]) *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. Assume the generalised Lindelöf hypothesis, and suppose that $R \asymp t_g^{-\delta}$ with $\delta < 1/3$. Then*

$$\frac{1}{\mathrm{vol}\,(B_R)}\int_{B_R(w)} |g(z)|^2\,d\mu(z) = \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1)$$

for every fixed point $w \in \Gamma\backslash\mathbb{H}$.

Similarly, let $g(z) = E(z, 1/2 + it_g)$, and suppose that $R \asymp t_g^{-\delta}$ with $\delta < 1/9$. Then unconditionally

$$\frac{1}{\log\left(\frac{1}{4} + t_g^2\right)\mathrm{vol}\,(B_R)}\int_{B_R(w)} |g(z)|^2\,d\mu(z) = \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1)$$

for every fixed point $w \in \Gamma\backslash\mathbb{H}$.

In fact, with little work, we can improve the range in Young's result for Eisenstein series.

Theorem 1.16 *Let $g(z) = E(z, 1/2 + it_g)$, and suppose that $R \asymp t_g^{-\delta}$ with $\delta < 1/6$. Then unconditionally*

$$\frac{1}{\log\left(\frac{1}{4} + t_g^2\right)\mathrm{vol}\,(B_R)}\int_{B_R(w)} |g(z)|^2\,d\mu(z) = \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1)$$

for every fixed point $w \in \Gamma\backslash\mathbb{H}$.

A simpler version of Question 1.12 is to instead consider eigenfunctions of the Laplacian on the d-torus $\mathbb{T}^d$ for any $d \geq 2$. Hezari and Rivière [23, Corollary 1.5] give strong bounds for equidistribution in shrinking balls along a full density subsequence of eigenfunctions of the Laplacian on $\mathbb{T}^d$ with eigenvalue λ, namely equidistribution on all balls of radius $R \gg \lambda^{-\frac{1}{4(d+1)}}$. Lester and Rudnick [35, Theorem 1.1] improve this to $R \gg_{\varepsilon} \lambda^{-\frac{1}{2(d-1)}+\varepsilon}$. Moreover, they prove [35, Theorems 3.1 and 4.1] that this is essentially sharp, in that there exists a subsequence of eigenfunctions for which equidistribution does not occur on shrinking balls of radius $R \ll_{\varepsilon} \lambda^{-\frac{1}{2(d-1)}-\varepsilon}$. For $d = 2$, Granville and Wigman [15, Corollary 3.2] have subsequently sharpened Lester and Rudnick's results to show there exists $A > 0$ such that equidistribution may not occur on shrinking balls of radius $R \ll_A \lambda^{-1/2}(\log \lambda)^A$.

One can also reformulate Question 1.12 probabilistically by asking for which scales equidistribution holds almost surely with respect to a random eigenbasis of Laplacian eigenfunctions; positive results towards this question appear in the work of Han [17] and Han and Tacy [18].

We study a related question: instead of demanding that equidistribution hold in shrinking balls of radius $R > 0$ centred at w for every point $w \in \Gamma\backslash\mathbb{H}$, we relax this requirement by instead asking whether equidistribution holds in shrinking balls $B_R(w)$ for almost every $w \in \Gamma\backslash\mathbb{H}$.

1.3.1 Conditional results

We are able to give a conditional proof of equidistribution in almost every shrinking ball when $g \in \mathcal{B}_0(\Gamma)$ and $R \gg t_g^{-\delta}$ for any $0 < \delta < 1$, that is, at all scales above the Planck scale.

Theorem 1.17 *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. Assume the generalised Lindelöf hypothesis, and suppose that $R \asymp t_g^{-\delta}$ for some $0 < \delta < 1$. Then for any $c \gg_{\varepsilon} t_g^{-\frac{1-\delta}{2}+\varepsilon}$,*

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) - \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}\right| > c\right\}\right)$$

converges to zero as t_g tends to infinity.

1.3.2 Unconditional results

Proving unconditional results seems to be much more difficult. Nevertheless, we are able to do so when $g(z) = E\left(z, 1/2 + it_g\right)$ is an Eisenstein series.

Theorem 1.18 *Let $g(z) = E\left(z, 1/2 + it_g\right)$. Suppose that $R \asymp t_g^{-\delta}$ for some $0 < \delta < 1$. Then for any $c \gg_{\varepsilon} t_g^{-\min\left\{\frac{5}{14}(1-\delta), 2\delta, \frac{1}{12}\right\}+\varepsilon}$,*

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) - D(g; w)\right| > c\right\}\right)$$

converges to zero as t_g tends to infinity, where $D(g; w)$ is given by (5.7).

This result is consistent with Theorem 1.6 due to the following.

Lemma 1.19 *In any compact subset K of $\Gamma\backslash\mathbb{H}$, we have that for all $w \in K$,*

$$D(g; w) = \frac{\log\left(\frac{1}{4} + t_g^2\right)}{\operatorname{vol}(\Gamma\backslash\mathbb{H})} + O_K\left(\left(\log t_g\right)^{2/3}\left(\log\log t_g\right)^{1/3}\right).$$

In particular, we may rephrase Theorem 1.18 in the following way.

Corollary 1.20 *Let $g(z) = E\left(z, 1/2 + it_g\right)$, and let K be a fixed compact subset of $\Gamma\backslash\mathbb{H}$. Suppose that $R \gg_\varepsilon t_g^{-1+\varepsilon}$. Then for any fixed $c > 0$,*

$$\operatorname{vol}\left(\left\{w \in K : \left|\frac{1}{\log\left(\frac{1}{4} + t_g^2\right)\operatorname{vol}(B_R)}\int_{B_R(w)} |g(z)|^2\, d\mu(z) - \frac{1}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\right| > c\right\}\right)$$

converges to zero as t_g tends to infinity.

1.4 Equidistribution of geometric invariants of quadratic fields in shrinking sets

Finally, in Sect. 6, we study a similar equidistribution problem in shrinking sets. Associated to each narrow ideal class A of the narrow class group Cl_K^+ of a quadratic number field $K = \mathbb{Q}(\sqrt{D})$ is a geometric invariant. For $D < 0$, this is a Heegner point z_A, while for $D > 0$, this is a closed geodesic $\mathcal{C}_A$ or a hyperbolic orbifold $\Gamma_A\backslash\mathcal{N}_A$ having this closed geodesic as its boundary; we explain these geometric invariants in more detail in Sect. 6.1.

For each fundamental discriminant D, we choose a genus $G_K \subset \mathrm{Cl}_K^+$ in the group of genera $\mathrm{Gen}_K = \mathrm{Cl}_K^+/(\mathrm{Cl}_K^+)^2$, so that G_K is a coset $A(\mathrm{Cl}_K^+)^2$ of narrow ideal classes in Cl_K^+. We have that $\mathrm{Gen}_K \cong (\mathbb{Z}/2\mathbb{Z})^{\omega(|D|)-1}$, where $\omega(|D|)$ is the number of distinct prime factors of $|D|$, so that $\#G_K = \#(\mathrm{Cl}_K^+)^2 = 2^{1-\omega(|D|)}h_K^+$, where $h_K^+ := \#\mathrm{Cl}_K^+$ denotes the narrow class number of K. Duke, Imamoḡlu, and Tóth have proved the following equidistribution theorem.

Theorem 1.21 ([11, Theorem 2]) *For every continuity set $B \subset \Gamma\backslash\mathbb{H}$,*

$$\frac{\#\{A \in G_K : z_A \in B\}}{\#G_K} = \frac{\operatorname{vol}(B)}{\operatorname{vol}(\Gamma\backslash\mathbb{H})} + o_B(1)$$

as $D \to -\infty$ through fundamental discriminants, and

$$\frac{\sum_{A\in G_K} \ell(\mathcal{C}_A \cap B)}{\sum_{A\in G_K} \ell(\mathcal{C}_A)} = \frac{\operatorname{vol}(B)}{\operatorname{vol}(\Gamma\backslash\mathbb{H})} + o_B(1),$$

$$\frac{\sum_{A\in G_K} \operatorname{vol}(\Gamma_A\backslash\mathcal{N}_A \cap B)}{\sum_{A\in G_K} \operatorname{vol}(\Gamma_A\backslash\mathcal{N}_A)} = \frac{\operatorname{vol}(B)}{\operatorname{vol}(\Gamma\backslash\mathbb{H})} + o_B(1)$$

as $D \to \infty$ through fundamental discriminants, where $\ell(\mathcal{C}_A) := \int_{\mathcal{C}_A} ds$, with $ds^2 = y^{-2}dx^2 + y^{-2}dy^2$.

If we sum over all genera, so that we are studying equidistribution associated to the full narrow class group, then this result is due to Duke [10, Theorem 1] for Heegner points and closed geodesics, while this result becomes trivial for hyperbolic orbifolds, for there is no error term whatsoever in this case. Moreover, the equidistribution of closed geodesics has a stronger realisation: instead of merely asking for the equidistribution of closed geodesics on $\Gamma\backslash\mathbb{H}$, we may lift these geodesics to phase space $S^*(\Gamma\backslash\mathbb{H}) \cong \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ and demand equidistribution with respect to the Liouville measure. This has been proved by Chelluri [8].

It is natural to ask whether equidistribution still occurs if B shrinks as $|D|$ grows. Towards this, Young [48] has proved the following.

Theorem 1.22 (Young [48, Theorem 2.1]) *Fix $w \in \Gamma\backslash\mathbb{H}$, and suppose that $R \asymp (-D)^{-\delta}$. Unconditionally, as $D \to -\infty$ through odd fundamental discriminants,*

$$\frac{\#\{A \in \mathrm{Cl}_K : z_A \in B_R(w)\}}{\mathrm{vol}(B_R)\, h_K} = \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1) \tag{1.23}$$

for fixed $\delta < 1/24$, where Cl_K denotes the class group of K and $h_K := \#\mathrm{Cl}_K$ denotes the class number. Assuming the generalised Lindelöf hypothesis, (1.23) holds as $D \to -\infty$ through fundamental discriminants for fixed $\delta < 1/8$.

In fact, from the method of proof, it is clear that Young's theorem applies to genera mutatis mutandis, and proves equidistribution not only of Heegner points, but also of closed geodesics and hyperbolic orbifolds.

Theorem 1.24 *Fix $w \in \Gamma\backslash\mathbb{H}$, and suppose that $R \asymp D^{-\delta}$. Unconditionally, as $D \to \infty$ through odd fundamental discriminants,*

$$\frac{\sum_{A\in G_K} \ell(\mathcal{C}_A \cap B_R(w))}{\mathrm{vol}(B_R)\sum_{A\in G_K}\ell(\mathcal{C}_A)} = \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1) \qquad \text{for } \delta < 1/18,$$
$$\frac{\sum_{A\in G_K} \mathrm{vol}(\Gamma_A\backslash\mathcal{N}_A \cap B_R(w))}{\mathrm{vol}(B_R)\sum_{A\in G_K}\mathrm{vol}(\Gamma_A\backslash\mathcal{N}_A)} = \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} + o_{w,\delta}(1) \qquad \text{for } \delta < 1/12.$$

Assuming the generalised Lindelöf hypothesis, these hold as $D \to \infty$ through fundamental discriminants for $\delta < 1/6$ and $\delta < 1/4$ respectively.

Once again, we may weaken the demand that equidistribution hold in shrinking balls of radius $R > 0$ centred at w for every point $w \in \Gamma\backslash\mathbb{H}$ and instead study whether equidistribution holds in shrinking balls $B_R(w)$ for almost every $w \in \Gamma\backslash\mathbb{H}$.

We prove the following conditional result.

Theorem 1.25 *Suppose that $R \asymp |D|^{-\delta}$. Assuming the generalised Lindelöf hypothesis, we have that for $0 < \delta < 1/4$ and $c \gg_\varepsilon (-D)^{-\frac{1}{2}\left(\frac{1}{4}-\delta\right)+\varepsilon}$,*

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{\#\{A \in G_K : z_A \in B_R(w)\}}{\mathrm{vol}(B_R)\,\#G_K} - \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})}\right| > c\right\}\right)$$

converges to zero as $D \to -\infty$ *along fundamental discriminants, while for* $0 < \delta < 1/2$ *and* $c \gg_{\varepsilon} D^{-\frac{1}{2}\left(\frac{1}{2}-\delta\right)+\varepsilon}$,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{\sum_{A\in G_K} \ell\left(\mathcal{C}_A \cap B_R(w)\right)}{\mathrm{vol}\left(B_R\right)\sum_{A\in G_K} \ell\left(\mathcal{C}_A\right)} - \frac{1}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right| > c\right\}\right)$$

converges to zero as $D \to \infty$ *along fundamental discriminants.*

Unconditionally, we obtain the following weaker results.

Theorem 1.26 *Suppose that* $R \asymp |D|^{-\delta}$. *We have that for* $0 < \delta < 1/12$ *and* $c \gg_{\varepsilon} (-D)^{-\frac{1}{2}\left(\frac{1}{12}-\delta\right)+\varepsilon}$,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{\#\left\{A \in G_K : z_A \in B_R(w)\right\}}{\mathrm{vol}\left(B_R\right)\#G_K} - \frac{1}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right| > c\right\}\right)$$

converges to zero as $D \to -\infty$ *along odd fundamental discriminants, while for* $0 < \delta < 1/6$ *and* $c \gg_{\varepsilon} D^{-\frac{1}{2}\left(\frac{1}{6}-\delta\right)+\varepsilon}$,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{\sum_{A\in G_K} \ell\left(\mathcal{C}_A \cap B_R(w)\right)}{\mathrm{vol}\left(B_R\right)\sum_{A\in G_K} \ell\left(\mathcal{C}_A\right)} - \frac{1}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right| > c\right\}\right)$$

converges to zero as $D \to \infty$ *along odd fundamental discriminants, and for all* $\delta > 0$ *and* $c \gg_{\varepsilon} D^{-1/4+\varepsilon}$,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{\sum_{A\in G_K} \mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A \cap B_R(w)\right)}{\mathrm{vol}\left(B_R\right)\sum_{A\in G_K} \mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right)} - \frac{1}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right| > c\right\}\right)$$

converges to zero as $D \to \infty$ *along odd fundamental discriminants.*

The fact that these geometric invariants equidistribute on almost every ball of different scales should not come as a surprise, and essentially boils down to the fact that a Heegner point has dimension 0, a closed geodesic has dimension 1, and a hyperbolic orbifold has dimension 2. For Heegner points, we need roughly R^2 balls to cover $\Gamma\backslash\mathbb{H}$, so we require the number of Heegner points $\#G_K$ corresponding to the genus G_K to be at least R^2 in order to expect equidistribution; this is the scale $R \asymp (-D)^{-1/4}$. For closed geodesics, on the other hand, R balls will cover roughly $1/R$ of $\Gamma\backslash\mathbb{H}$, but a closed geodesic may intersect more than one ball, so we only require the total length $\sum_{A\in G_K} \ell\left(\mathcal{C}_A\right)$ of closed geodesics corresponding to the genus G_K to be at least R; this is the scale $R \asymp D^{-1/2}$. Finally, we should expect equidistribution at *all* scales for hyperbolic orbifolds, since these are just (possibly uneven) coverings of $\Gamma\backslash\mathbb{H}$.

1.5 Idea of proof

The chief idea behind the proof of the aforementioned small scall equidistribution theorems is to use Chebyshev's inequality to reduce the problem to bounding a variance. For example,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) - \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}\right| > c\right\}\right) \leq \frac{1}{c^2} \mathrm{Var}(g; R)$$

with

$$\mathrm{Var}(g; R) := \int_{\Gamma\backslash\mathbb{H}} \left(\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) - \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}\right)^2 d\mu(w).$$

The method of bounding the variance in order to show equidistribution in almost every shrinking ball is also used in [15, Theorem 1.6] for eigenfunctions of the Laplacian on $\mathbb{T}^2$, as well as in both [13, Theorem 1.3] and [4, Theorem 1.8], where the problem investigated is not quantum unique ergodicity, but rather the equidistribution of lattice points on the sphere.

The variance is an inner product of functions in $L^2(\Gamma\backslash\mathbb{H})$, as is the fourth moment of a truncated Eisenstein series; both are thereby amenable to being spectrally expanded via Parseval's identity. The resulting spectral sum over Hecke–Maaß forms f occurring in the spectral expansion $\mathrm{Var}(g; R)$ when g is an Eisenstein series is essentially the same as the spectral sum for fourth moment of a truncated Eisenstein series in the range $0 < t_f \ll_\varepsilon R^{-1+\varepsilon}$, whereas for $t_f \gg 1/R$, it is much smaller.

Finally, we use the Watson–Ichino formula to write $|\langle |g|^2, f\rangle|^2$ as a product of L-functions. This reduces the problem to bounding certain moments of L-functions, with the length of these moments corresponding inversely to the radius of the shrinking ball.

Though not a manifestation of the random wave conjecture, the equidistribution problems in Sect. 1.4 nonetheless involve equidistribution on $\Gamma\backslash\mathbb{H}$, and the proofs of Theorems 1.25 and 1.26 contain many of the same ingredients as the proofs of Theorems 1.17 and 1.18. The chief difference is that in place of $|\langle |g|^2, f\rangle|^2$, we have Weyl sums; akin to the Watson–Ichino formula, these can be expressed as a product of L-functions via the work of Duke, Imamoḡlu, and Tóth [11].

1.6 Connections to subconvexity

The rate of equidistribution for quantum unique ergodicity for Hecke–Maaß eigenforms $g \in \mathcal{B}_0(\Gamma)$ can be quantified via explicit rates of decay for

$$\int_{\Gamma\backslash\mathbb{H}} f(z)|g(z)|^2 \, d\mu(z), \qquad \int_{\Gamma\backslash\mathbb{H}} E(z, \psi)|g(z)|^2 \, d\mu(z)$$

for fixed $f \in \mathcal{B}_0(\Gamma)$ and $\psi \in C_c^\infty(\mathbb{R}^+)$ as t_g tends to infinity. Via the Watson–Ichino formula, this is equivalent to obtaining subconvex bounds of the form

$$L\left(\frac{1}{2}, \operatorname{sym}^2 g \otimes f\right) \ll_f t_g^{1-\delta}, \qquad L\left(\frac{1}{2} + it, \operatorname{sym}^2 g\right) \ll_t t_g^{\frac{1}{2}(1-\delta)}$$

for some absolute constant $\delta > 0$. Similarly, quantifying the rate of equidistribution for quantum unique ergodicity for $g(z) = E(z, 1/2 + it_g)$ is equivalent to obtaining subconvex bounds of the form

$$L\left(\frac{1}{2} + 2it_g, f\right) \ll_f t_g^{\frac{1}{2}(1-\delta)}, \qquad \zeta\left(\frac{1}{2} + i(2t_g \pm t)\right) \ll_t t_g^{\frac{1}{4}(1-\delta)}$$

for some absolute constant $\delta > 0$.

For quantum unique ergodicity in almost every shrinking ball of radius R for Hecke–Maaß eigenforms $g \in \mathcal{B}_0(\Gamma)$, on the other hand, we will show that we require bounds of the form

$$\sum_{H \leq t_f \leq 2H} \frac{L\left(\frac{1}{2}, f\right) L\left(\frac{1}{2}, \operatorname{sym}^2 g \otimes f\right)}{L(1, \operatorname{sym}^2 f)} \ll_\delta H t_g^{1-\delta}$$

for some absolute constant $\delta > 0$ *uniformly in* $1 \ll H \ll 1/R$. That is, we require subconvex moment bounds for L-functions uniformly in two parameters: t_f and t_g. Thus this is a problem of hybrid subconvexity. Proving such bounds unconditionally seems to be currently out of reach for moments involving $\mathrm{GL}_3 \times \mathrm{GL}_2$ Rankin–Selberg L-functions. For $g(z) = E(z, 1/2 + it_g)$, on the other hand, the required subconvex moment bounds are

$$\sum_{H \leq t_f \leq 2H} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2} + 2it_g\right)\right|^2}{L(1, \operatorname{sym}^2 f)} \ll_\delta H t_g^{1-\delta},$$

and the fact that these moments only involve GL_2 L-functions makes this problem tractable. It is for this reason that we are able to prove Theorem 1.18 unconditionally, whereas Theorem 1.17 is conditional.

2 Integrals of automorphic forms and L-functions

2.1 The Maaß–Selberg relation

The Eisenstein series $E(z, 1/2 + it)$ is not square-integrable for any $t \in \mathbb{R}$. However, this is no longer the case when we replace the Eisenstein series with the truncated Eisenstein series

$$g(z) = \Lambda^T E\left(z, \frac{1}{2} + it_g\right),$$

since $\Lambda^T E(z,s)$ is of rapid decay at the cusp of $\Gamma\backslash\mathbb{H}$. Note that

$$\Lambda^T E(z,s) = \begin{cases} E(z,s) & \text{if } 1/T \leq \Im(z) \leq T, \\ E(z,s) - \Im(z)^s + \varphi(s)\Im(z)^{1-s} & \text{if } \Im(z) > T, \end{cases}$$

where

$$\varphi(s) = \frac{\Lambda(2-2s)}{\Lambda(2s)}.$$

The following explicit formula for the inner product of two truncated Eisenstein series is known as the Maaß–Selberg relation.

Proposition 2.1 ([28, Proposition 6.8]) *For $T \geq 1$, and $s \neq \overline{r}$, $s + \overline{r} \neq 1$,*

$$\int_{\Gamma\backslash\mathbb{H}} \Lambda^T E(z,s)\overline{\Lambda^T E(z,r)}\, d\mu(z) = \frac{T^{s+\overline{r}-1}}{s+\overline{r}-1} + \overline{\varphi(r)}\frac{T^{s-\overline{r}}}{s-\overline{r}} + \varphi(s)\frac{T^{\overline{r}-s}}{\overline{r}-s} + \varphi(s)\overline{\varphi(r)}\frac{T^{1-s-\overline{r}}}{1-s-\overline{r}}. \tag{2.2}$$

Corollary 2.3 *We have that*

$$\int_{\Gamma\backslash\mathbb{H}} \left|\Lambda^T E\left(z, \frac{1}{2} + it_g\right)\right|^2 d\mu(z) = \log\left(\left(\frac{1}{4} + t_g^2\right) T^2\right) + O\left((\log t_g)^{2/3} (\log\log t_g)^{1/3}\right).$$

Proof We take $s = r = 1/2 + it_g + \varepsilon$ with $\varepsilon > 0$ in the Maaß–Selberg relation (2.2) to obtain

$$\int_{\Gamma\backslash\mathbb{H}} \left|\Lambda^T E\left(z, \frac{1}{2} + it_g + \varepsilon\right)\right|^2 d\mu(z) = \frac{T^{2\varepsilon}}{2\varepsilon} - \left|\varphi\left(\frac{1}{2} + it_g + \varepsilon\right)\right|^2 \frac{T^{-2\varepsilon}}{2\varepsilon}.$$

Using the Taylor expansions

$$T^{2\varepsilon} = 1 + 2\varepsilon \log T + O\left(\varepsilon^2\right),$$
$$\varphi\left(\frac{1}{2} + it_g + \varepsilon\right) = \varphi\left(\frac{1}{2} + it_g\right) + \varepsilon\varphi'\left(\frac{1}{2} + it_g\right) + O\left(\varepsilon^2\right),$$

together with the fact that $|\varphi(1/2 + it_g)| = 1$ and that

$$\frac{\varphi'}{\varphi}\left(\frac{1}{2} + it_g\right) = -4\Re\left(\frac{\Lambda'}{\Lambda}\left(1 + 2it_g\right)\right) = 2\log\pi - 2\Re\left(\frac{\Gamma'}{\Gamma}\left(\frac{1}{2} + it_g\right)\right) - 4\Re\left(\frac{\zeta'}{\zeta}\left(1 + 2it_g\right)\right), \tag{2.4}$$

we find that

$$\int_{\Gamma\backslash\mathbb{H}} |g(z)|^2\, d\mu(z) = 2\log T - 2\log\pi + 2\Re\left(\frac{\Gamma'}{\Gamma}\left(\frac{1}{2}+it_g\right)\right) + 4\Re\left(\frac{\zeta'}{\zeta}\left(1+2it_g\right)\right).$$

It remains to use Stirling's formula to find that

$$2\Re\left(\frac{\Gamma'}{\Gamma}\left(\frac{1}{2}+it_g\right)\right) = \log\left(\frac{1}{4}+t_g^2\right) + O\left(\frac{1}{t_g}\right), \tag{2.5}$$

and [29, Theorem 8.29] to give the bound

$$\frac{\zeta'}{\zeta}\left(1+2it_g\right) \ll \left(\log t_g\right)^{2/3}\left(\log\log t_g\right)^{1/3}. \tag{2.6}$$

□

2.2 The Watson–Ichino formula

To deal with spectral sums involving terms of the form $|\langle |g|^2, f\rangle|^2$, one can use the Watson–Ichino formula, which essentially states that the square of the integral over $\Gamma\backslash\mathbb{H}$ of the product of three automorphic forms is equal to a product of completed L-functions involving those automorphic forms. In particular, if $f, g \in \mathcal{B}_0(\Gamma)$, then from [26, Theorem 1.1] and [46, Theorem 3],

$$\left|\left\langle |g|^2, f\right\rangle\right|^2 = \frac{\Lambda\left(\frac{1}{2}, g\otimes\widetilde{g}\otimes f\right)}{\Lambda(1,\operatorname{sym}^2 g)^2\Lambda\left(1,\operatorname{sym}^2 f\right)}.$$

Here $\Lambda(s,\pi)$ denotes the completed L-function of an automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_\mathbb{Q})$: this is of the form

$$\Lambda(s,\pi) = q_\pi^{s/2} L_\infty(s,\pi) L(s,\pi), \tag{2.7}$$

where q_π denotes the conductor of π, $L_\infty(s,\pi)$ is the archimedean part of $\Lambda(s,\pi)$, which is of the form $\pi^{-ns/2}\prod_{j=1}^n \Gamma(\frac{s+\kappa_{\pi,j}}{2})$ for some $\kappa_{\pi,j}\in\mathbb{C}$, and $L(s,\pi)$ is the usual nonarchimedean part of $\Lambda(s,\pi)$. Note that the numerator in the Watson–Ichino formula factorises:

$$\Lambda\left(s, g\otimes\widetilde{g}\otimes f\right) = \Lambda(s,f)\Lambda\left(s,\operatorname{sym}^2 g\otimes f\right).$$

Similar results also hold when either f or g is replaced with an Eisenstein series.

Proposition 2.8 ([6, Equations (2.2) and (4.2)]) *For* $f, g\in\mathcal{B}_0(\Gamma)$,

$$\left|\left\langle |g|^2, f\right\rangle\right|^2 = \frac{1}{8}\frac{\Lambda\left(\frac{1}{2},f\right)\Lambda\left(\frac{1}{2},\operatorname{sym}^2 g\otimes f\right)}{\Lambda\left(1,\operatorname{sym}^2 g\right)^2\Lambda\left(1,\operatorname{sym}^2 f\right)},$$

$$\left|\left\langle |g|^2, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle\right|^2 = \frac{1}{4}\frac{\Lambda\left(\frac{1}{2}+it\right)\Lambda\left(\frac{1}{2}-it\right)\Lambda\left(\frac{1}{2}+it, \mathrm{sym}^2 g\right)\Lambda\left(\frac{1}{2}-it, \mathrm{sym}^2 g\right)}{\Lambda\left(1, \mathrm{sym}^2 g\right)^2 \Lambda(1+2it)\Lambda(1-2it)}.$$

A similar result also holds when g is an Eisenstein series.

Proposition 2.9 ([39, Equation (17)], [45, Theorem 4.1]) *For* $f \in \mathcal{B}_0(\Gamma)$,

$$\left|\left\langle \left|E\left(\cdot, \frac{1}{2}+it\right)\right|^2, f\right\rangle\right|^2 = \frac{1}{2}\frac{\Lambda\left(\frac{1}{2}, f\right)^2 \Lambda\left(\frac{1}{2}+2it_g, f\right)\Lambda\left(\frac{1}{2}-2it_g, f\right)}{\Lambda(1+2it_g)^2\Lambda\left(1-2it_g\right)^2\Lambda\left(1, \mathrm{sym}^2 f\right)}.$$

Finally, when f is also an Eisenstein series, the integral is no longer convergent. One can work around this issue by replacing this integral with a regularised integral. This is defined by Zagier [50] in the following way. Let $F : \Gamma\backslash\mathbb{H} \to \mathbb{C}$ be a continuous function of moderate growth, so that there exists $c_j, \alpha_j \in \mathbb{C}$ and nonnegative integers n_j such that

$$F(z) = \sum_{j=1}^{\ell} \frac{c_j}{n_j!} y^{\alpha_j} (\log y)^{n_j} + O_N\left(y^{-N}\right)$$

for all $N \geq 0$ at the cusp at infinity, with no α_j equal to 0 or 1. Then there exists a function $\mathcal{E}(z)$ that is a linear combination of Eisenstein series and derivatives of Eisenstein series $E(z, \alpha)$, each satisfying $\Re(\alpha) > 1/2$, such that for some $\delta > 0$,

$$F(z) - \mathcal{E}(z) = O\left(y^{\frac{1}{2}-\delta}\right)$$

at the cusp at infinity. The regularised inner product of two functions f, g such that $f\overline{g} = F$ is continuous and of moderate growth is defined to be

$$\langle f, g\rangle_{\mathrm{reg}} := \int_{\Gamma\backslash\mathbb{H}} (F(z) - \mathcal{E}(z))\, d\mu(z).$$

Moreover, if f and g depend on complex parameters, then we may extend both sides via analytic continuation where possible.

Proposition 2.10 ([50, Equation (44)]) *We have that*

$$\langle E(\cdot, s_1)E(\cdot, s_2), E(\cdot, s)\rangle_{\mathrm{reg}} = \frac{\Lambda(\overline{s}+s_1+s_2-1)\Lambda(\overline{s}+s_1-s_2)\Lambda(\overline{s}-s_1+s_2)\Lambda(\overline{s}-s_1-s_2+1)}{\Lambda(2\overline{s})\Lambda(2s_1)\Lambda(2s_2)}. \tag{2.11}$$

In practice, it is the nonarchimedean part $L(s, \pi)$ of a completed L-function $\Lambda(s, \pi)$ that is difficult to deal with; this is because the asymptotic behaviour of the archimedean part of a completed L-function can be inferred via Stirling's approximation.

Lemma 2.12 *The product of the archimedean parts of the completed L-functions in Propositions 2.8, (2.9) (with $t = t_f$), and (2.10) (with $s_1 = s_2 = 1/2 + it_g$ and $s = 1/2 + it_f$) is equal to*

$$\frac{8\pi^2 e^{-\pi\Omega(t_f,t_g)}}{(1+t_f)(1+2t_g+t_f)^{1/2}(1+|2t_g-t_f|)^{1/2}} \times\left(1+O\left(\frac{1}{1+t_f}+\frac{1}{1+2t_g+t_f}+\frac{1}{1+|2t_g-t_f|}\right)\right), \tag{2.13}$$

where

$$\Omega(t_f,t_g) := \begin{cases} 0 & \text{if } 0 < t_f \le 2t_g, \\ t_f - 2t_g & \text{if } t_f > 2t_g. \end{cases}$$

Proof The product of the archimedean parts of the completed L-functions is

$$\pi\frac{\Gamma\left(\frac{1}{4}+\frac{i(2t_g+t_f)}{2}\right)\Gamma\left(\frac{1}{4}+\frac{i(2t_g-t_f)}{2}\right)\Gamma\left(\frac{1}{4}-\frac{i(2t_g+t_f)}{2}\right)\Gamma\left(\frac{1}{4}-\frac{i(2t_g-t_f)}{2}\right)}{\Gamma\left(\frac{1}{2}+it_g\right)^2\Gamma\left(\frac{1}{2}-it_g\right)^2} \times\frac{\Gamma\left(\frac{1}{4}+\frac{it_f}{2}\right)^2\Gamma\left(\frac{1}{4}-\frac{it_f}{2}\right)^2}{\Gamma\left(\frac{1}{2}+it_f\right)\Gamma\left(\frac{1}{2}-it_f\right)}.$$

The result then follows directly from Stirling's approximation. $\square$

On occasion, we also need to deal with lower bounds for $L(1, \mathrm{sym}^2 f)$. This is less complex than values of L-functions within the critical strip $0 < \Re(s) < 1$; indeed, the following is known.

Lemma 2.14 (Hoffstein–Lockhart [24]) *For $f \in \mathcal{B}_0(\Gamma)$,*

$$L\left(1, \mathrm{sym}^2 f\right) \gg \frac{1}{\log(t_f+3)}.$$

3 Sharp bounds for the L^4-norm of a truncated Eisenstein series

3.1 The spectral expansion of the L^4-norm

We wish to determine sharp bounds for

$$\|g\|_{L^4(\Gamma\backslash\mathbb{H})}^4 = \int_{\Gamma\backslash\mathbb{H}} |g(z)|^4\, d\mu(z)$$

with $g(z) = \Lambda^T E(z, 1/2+it_g)$ in terms of t_g. Our first step is to express this quantity as a spectral sum, which requires the spectral decomposition of $L^2(\Gamma\backslash\mathbb{H})$.

Lemma 3.1 ([29, Theorem 15.5]) *Let*

$$f_0(z) := \frac{1}{\sqrt{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}},$$

so that $\langle f_0, f_0\rangle = 1$*, and let* $\mathcal{B}_0(\Gamma)$ *be an orthonormal basis of Maaß cusp forms in* $L^2\,(\Gamma\backslash\mathbb{H})$*. Then a function* $g \in L^2\,(\Gamma\backslash\mathbb{H})$ *has the spectral expansion, valid in the* L^2*-sense, of the form*

$$g(z) = \langle g, f_0\rangle\, f_0(z) + \sum_{f\in\mathcal{B}_0(\Gamma))} \langle g, f\rangle\, f(z)$$
$$+\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\langle g, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle E\left(z, \frac{1}{2}+it\right)\, dt.$$

Moreover, Parseval's identity holds:

$$\langle g_1, g_2\rangle = \langle g_1, f_0\rangle\,\langle f_0, g_2\rangle + \sum_{f\in\mathcal{B}_0(\Gamma)} \langle g_1, f\rangle\,\langle f, g_2\rangle$$
$$+\frac{1}{4\pi}\int_{-\infty}^{\infty}\left\langle g_1, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle\left\langle E\left(\cdot, \frac{1}{2}+it\right), g_2\right\rangle dt$$

for $g_1, g_2 \in L^2\,(\Gamma\backslash\mathbb{H})$.

In particular, the following spectral expansion of the L^4-norm of g is simply Parseval's identity with $g_1 = g_2 = |g|^2$.

Corollary 3.2 *Let* $g \in L^2\,(\Gamma\backslash\mathbb{H})$ *be of rapid decay. Then*

$$\|g\|_{L^4(\Gamma\backslash\mathbb{H})}^4 = \left|\left\langle |g|^2, f_0\right\rangle\right|^2 + \sum_{f\in\mathcal{B}_0(\Gamma)} \left|\left\langle |g|^2, f\right\rangle\right|^2 + \frac{1}{4\pi}\int_{-\infty}^{\infty}\left|\left\langle |g|^2, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle\right|^2 dt.$$

This is reduced to understanding bounds for the inner product of $|g|^2$ with eigenfunctions of the Laplacian. The first term in this expansion is the inner product of $|g|^2$ with the constant function

$$f_0(z) = \frac{1}{\sqrt{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}},$$

and Corollary 2.3 shows that

$$\left|\left\langle |g|^2, f_0\right\rangle\right|^2 = \frac{\left(\log\left(\frac{1}{4}+t_g^2\right)\right)^2}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} + O_T\left(\left(\log t_g\right)^{5/3}\left(\log\log t_g\right)^{1/3}\right).$$

It remains to treat the cuspidal and continuous spectra.

3.2 Ranges of the spectral decomposition for the L^4-norm

We divide the spectral expansion of the L^4-norm of $g(z) = \Lambda^T E(z, 1/2 + it_g)$ given in Corollary 3.2 into different parts, then analyse each part individually.

There are two main ranges of the continuous spectrum to consider, which depend on a small fixed parameter $\delta > 0$:

- the initial range $0 \leq |t| \leq 2t_g + t_g^{1-\delta}$, and
- the tail range $|t| > 2t_g + t_g^{1-\delta}$.

Both of these ranges will be shown to contribute a negligible amount via subconvexity estimates for the L-functions appearing in the integral.

For the contribution from the cuspidal spectrum, the summation over $\mathcal{B}_0(\Gamma)$ may be broken up into different ranges depending on t_f. There are four main ranges of the cuspidal spectrum left to consider, which depend on a fixed small parameter $\delta > 0$:

- the short initial range $0 \leq t_f \leq t_g^{1-\delta}$,
- the bulk range $t_g^{1-\delta} < t_f < 2t_g - t_g^{1-\delta}$,
- the short transition range $2t_g - t_g^{1-\delta} \leq t_f \leq 2t_g + t_g^{1-\delta}$, and
- the tail range $t_f > 2t_g + t_g^{1-\delta}$.

We divide the spectral sum into these particular ranges due to the size of the product of analytic conductors of L-functions. The analytic conductor of

$$L\left(\frac{1}{2}, f\right)^2 L\left(\frac{1}{2} + 2it_g, f\right) L\left(\frac{1}{2} - 2it_g, f\right)$$

is approximately

$$\left(\frac{1}{4} + t_f^2\right)^2 \left(\frac{1}{4} + \left(2t_g^2 + t_f^2\right)\right) \left(\frac{1}{4} + \left|2t_g^2 - t_f^2\right|\right),$$

which is large when t_f lies in the bulk range, but is small in the short initial range, and drops in the short transition range. For this reason, the main contribution will be shown to arise from the bulk range, while the contribution from the two short ranges will be shown to be negligible. Assuming the generalised Lindelöf hypothesis, this can be proven directly; see [6, Sect. 5]. Finally, the exponential decay in (2.13) arising from the archimedean components of the completed L-functions indicates that the tail range contributes a negligible amount.

Remark 3.3 In [45, Chapter 6], Spinu sketches an unconditional proof of Theorem 1.9. The proof, however, only treats the spectral sum in the range $\alpha t_g < t_f < 2(1-\alpha)t_g$ for any fixed $\alpha > 0$ (essentially the bulk range), in which the contribution of the spectral sum ought to be nonnegligible. The remaining ranges, which all ought to contribute a negligible amount, are left unaddressed.

This same issue is present in a claim of Sarnak and Watson [43, Theorem 3(a)] of the bound $\|g\|_{L^4(\Gamma\backslash\mathbb{H})} \ll_\varepsilon t_g^\varepsilon$ for Hecke–Maaß cusp forms, under the assumption of the Selberg eigenvalue and Ramanujan conjectures (but not the generalised Lindelöf hypothesis, as in [6, Theorem 1.1]). Sarnak (personal communication) subsequently has retracted this claim, and instead only claims this bound for the contribution of the spectral sum in the bulk range, as the method he uses is unable to treat the short initial range.

We are able to treat the short initial and transition ranges, left untreated by Spinu, by applying the work of Jutila [32], Ivić [27], and Jutila and Motohashi [33] on certain hybrid moments of L-functions. We do not know how to treat these ranges when g is a Hecke–Maaß cusp form.

3.3 Spectral methods to bound the continuous spectrum

From Corollary 3.2, we must bound

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\left|\left\langle |g|^2, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle\right|^2 dt. \tag{3.4}$$

Lemma 3.5 ([45, Theorem 3.3]) *There exists a positive constant* $c>0$ *such that* (3.4) *is bounded by* $108T + O\left(t_g^{-c}\right)$.

Here c is any constant less than $1/2-2\theta$, where θ is a positive constant such that

$$\zeta\left(\frac{1}{2}+it\right) \ll_{\varepsilon} (|t|+1)^{\theta+\varepsilon}.$$

The best bound known is $\theta = 13/84$, due to Bourgain [3, Theorem 5].

3.4 Reduction to untruncated Eisenstein series for the cuspidal spectrum

From Corollary 3.2, we must bound

$$\sum_{f\in\mathcal{B}_0(\Gamma)}\left|\left\langle |g|^2, f\right\rangle\right|^2.$$

First, we observe that $g(z) = \Lambda^T E\left(z, 1/2+it_g\right)$ can be replaced by $E\left(z, 1/2+it_g\right)$.

Lemma 3.6 ([45, Theorem 4.2]) *We have that*

$$\sum_{f\in\mathcal{B}_0(\Gamma)}\left|\left\langle |g|^2, f\right\rangle\right|^2 \leq \sum_{f\in\mathcal{B}_0(\Gamma)}\left|\left\langle\left|E\left(\cdot, \frac{1}{2}+it_g\right)\right|^2, f\right\rangle\right|^2 + O_T\left(\left(\log t_g\right)^2\right).$$

This allows us to use Proposition 2.9 and Lemma 2.12. We divide the cuspidal spectrum into four ranges, as discussed in Sect. 3.2. The convexity bound for the associated L-functions together with the Weyl law shows that the tail range is negligible. So it remains to bound the first three ranges.

3.5 Weaker bounds via the large sieve

In [45, Chapter 5], Spinu proves the following moment bounds in dyadic intervals, a corollary of which is the bound $\|g\|_{L^4(\Gamma\backslash\mathbb{H})} \ll_\varepsilon t_g^\varepsilon$.

Lemma 3.7 ([45, Proposition 5.4]) *We have that*

$$\sum_{H \leq t_f \leq 2H} L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2 \ll_\varepsilon Ht_g^{1+\varepsilon}$$

uniformly in $H \leq 2t_g - t_g^{1-\delta}$.

Lemma 3.8 ([45, Proposition 5.5]) *We have that*

$$\sum_{H<|t_f-2t_g|<2H} L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2 \ll H^{1/2} t_g^{\frac{3+\delta}{2}}$$

uniformly in $1 \leq H \ll t_g^{1-\delta}$.

Remark 3.9 Spinu uses the large sieve only to prove Lemma 3.7 and employs a more complex method in proving Lemma 3.8; nonetheless, one can in fact use the local large sieve, as stated in [38, Lemma], to prove the latter; see [38, Proof of Theorem].

3.6 Spectral methods to bound the short initial range

From [29, Theorem 8.29], we have that bound

$$\frac{1}{\zeta(1+it)} \ll (\log t)^{2/3} (\log\log t)^{1/3}.$$

It therefore suffices to show that

$$\sum_{0<t_f<t_g^{1-\delta}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{\left(1+t_f\right)\left(1+2t_g+t_f\right)^{1/2}\left(1+2t_g-t_f\right)^{1/2} L\left(1, \operatorname{sym}^2 f\right)} \ll t_g^{-\delta'}$$

for some $\delta' > 0$. We divide the short transition range $0 < t_f < t_g^{1-\delta}$ into dyadic intervals $H \leq t_f < 2H$, of which there are roughly $\log t_g$ intervals, on which

$$(1+t_f)(1+2t_g+t_f)^{1/2}(1+2t_g-t_f)^{1/2} \asymp Ht_g.$$

It then suffices to show that for $H \ll t_g^{1-\delta}$,

$$\sum_{H \leq t_f \leq 2H} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{L\left(1, \operatorname{sym}^2 f\right)} \ll Ht_g^{1-\delta'}.$$

This bound follows from the work of Jutila [32], Ivić [27], and Jutila and Motohashi [33]. It is worth noting that the purpose of these works is to obtain Weyl-type subconvexity bounds

$$L\left(\frac{1}{2}+it,f\right) \ll_{\varepsilon} \mathfrak{q}\left(f,\frac{1}{2}+it\right)^{\frac{1}{6}+\varepsilon}$$

for Hecke–Maaß eigenforms $f \in \mathcal{B}_0(\Gamma)$, so long as $|t|$ is not too close to t_f; here $\mathfrak{q}(f,s)$ denotes the analytic conductor of $L(s,f)$. Conveniently, their methods to obtain such bounds involve obtaining bounds for the exact type of spectral sum that we are studying.

Lemma 3.10 *For $t \geq 0$ and $H \gg 1$, we have that*

$$\sum_{H \leq t_f \leq 2H} \frac{L\left(\frac{1}{2},f\right)^2 \left|L\left(\frac{1}{2}+it,f\right)\right|^2}{L\left(1,\mathrm{sym}^2 f\right)} \ll_{\varepsilon} \begin{cases} H^{2+\varepsilon} & \text{if } H \geq t^{2/3}, \\ t^{\frac{4}{3}+\varepsilon} & \text{if } t^{1/2} \leq H \leq t^{2/3}, \\ H^{\frac{8}{3}+\varepsilon} & \text{if } t^{1/3} \leq H \leq t^{1/2}, \\ H^{\frac{2}{3}+\varepsilon} t^{\frac{2}{3}+\varepsilon} & \text{if } H \leq t^{1/3}. \end{cases}$$

Proof For $H \geq t^{1/2}$, this follows from [33, Theorem 2], which states that for $t \geq 0$ and $H \gg 1$,

$$\sum_{H \leq t_f \leq 2H} \frac{L\left(\frac{1}{2},f\right)^2 \left|L\left(\frac{1}{2}+it,f\right)\right|^2}{L\left(1,\mathrm{sym}^2 f\right)} \ll_{\varepsilon} \left(H^2+t^{4/3}\right)^{1+\varepsilon}.$$

For $H \leq t^{1/2}$, this follows from the subconvexity bound

$$L\left(\frac{1}{2},f\right) \ll_{\varepsilon} t_f^{\frac{1}{3}+\varepsilon}$$

of Ivić [27, Corollary 2], and from [32, Theorem], which states that for $t \geq 0$ and $1 \ll G \ll H$,

$$\sum_{H \leq t_f \leq H+G} \frac{\left|L\left(\frac{1}{2}+it,f\right)\right|^2}{L\left(1,\mathrm{sym}^2 f\right)} \ll_{\varepsilon} \left(GH+t^{2/3}\right)^{1+\varepsilon}.$$

□

Corollary 3.11 *For any $\delta > 0$, we have that*

$$\sum_{0 < t_f < t_g^{1-\delta}} \frac{\Lambda\left(\frac{1}{2},f\right)^2 \Lambda\left(\frac{1}{2}+2it_g,f\right) \Lambda\left(\frac{1}{2}-2it_g,f\right)}{\Lambda(1+2it_g)^2 \Lambda(1-2it_g)^2 \Lambda\left(1,\mathrm{sym}^2 f\right)} \ll_{\varepsilon} t_g^{-\min\left\{\delta,\frac{1}{6}\right\}+\varepsilon}.$$

3.7 Spectral methods to bound the short transition range

In [5, Sect. 1], Buttcane and Khan state for a dihedral Maaß newform g,

> …the range $[2t_g - t_g^{1-\delta} < t_f < 2t_g]$ can be handled by applying Hölder's inequality as Luo does and then applying Jutila's [31] and Ivić's [27] bounds for moments of $L(1/2, f)$ in short intervals of t_f close to $2t_g$.

A similar idea works when g is a truncated Eisenstein series. We must show that

$$\sum_{2t_g - t_g^{1-\delta} \leq t_f \leq 2t_g + t_g^{1-\delta}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2} + 2it_g, f\right)\right|^2}{(1+t_f)\left(1 + 2t_g + t_f\right)^{1/2} \left(1 + |2t_g - t_f|\right)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \ll t_g^{-\delta'}$$

for some $\delta' > 0$. We use the Cauchy–Schwarz inequality to see that this spectral sum is bounded by $t_g^{-3/2}$ times the square root of the product of

$$\sum_{2t_g - t_g^{1-\delta} \leq t_f \leq 2t_g + t_g^{1-\delta}} \frac{L\left(\frac{1}{2}, f\right)^4}{\left(1 + |2t_g - t_f|\right)^{1/2} L\left(1, \mathrm{sym}^2 f\right)}$$

and

$$\sum_{2t_g - t_g^{1-\delta} \leq t_f \leq 2t_g + t_g^{1-\delta}} \frac{\left|L\left(\frac{1}{2} + 2it_g, f\right)\right|^4}{\left(1 + |2t_g - t_f|\right)^{1/2} L\left(1, \mathrm{sym}^2 f\right)}.$$

The first sum is bounded by

$$\sum_{k=0}^{\lfloor t_g^{2/3-\delta} \rfloor} \frac{1}{\left(1 + kt_g^{1/3}\right)^{1/2}} \sum_{2t_g - (k+1)t_g^{1/3} \leq t_f < 2t_g - kt_g^{1/3}} \frac{L\left(\frac{1}{2}, f\right)^4}{L\left(1, \mathrm{sym}^2 f\right)}$$
$$+ \sum_{k=0}^{\lfloor t_g^{2/3-\delta} \rfloor} \frac{1}{\left(1 + kt_g^{1/3}\right)^{1/2}} \sum_{2t_g + kt_g^{1/3} \leq t_f < 2t_g + (k+1)t_g^{1/3}} \frac{L\left(\frac{1}{2}, f\right)^4}{L\left(1, \mathrm{sym}^2 f\right)},$$

and a similar expression holds for the second sum. We then apply the following lemma to show that each sum is bounded by a constant multiple dependent on ε of $t_g^{\frac{3-\delta}{2}+\varepsilon}$, from which the result follows.

Lemma 3.12 ([31, Theorem], [33, Theorem 1]) *For $H \gg 1$ and $1 \ll G \ll H$, we have that*

$$\sum_{H \leq t_f \leq H+G} \frac{L\left(\frac{1}{2}, f\right)^4}{L\left(1, \mathrm{sym}^2 f\right)} \ll_\varepsilon \left(H^{1/3} + G\right) H^{1+\varepsilon}.$$

Similarly, for $H \gg 1$, $0 \leq t \ll H^{3/2-\varepsilon}$, *and* $0 \leq G \leq (H+t)^{4/3} H^{-1+\varepsilon}$, *we have that*

$$\sum_{H \leq t_f \leq H+G} \frac{\left|L\left(\frac{1}{2}+it, f\right)\right|^4}{L\left(1, \mathrm{sym}^2 f\right)} \ll_\varepsilon (H+t)^{4/3} H^{\varepsilon}.$$

Corollary 3.13 *For any* $0 < \delta < 2/3$, *we have that*

$$\sum_{2t_g - t_g^{1-\delta} \leq t_f \leq 2t_g + t_g^{1-\delta}} \frac{\Lambda\left(\frac{1}{2}, f\right)^2 \Lambda\left(\frac{1}{2}+2it_g, f\right) \Lambda\left(\frac{1}{2}-2it_g, f\right)}{\Lambda(1+2it_g)^2 \Lambda(1-2it_g)^2 \Lambda\left(1, \mathrm{sym}^2 f\right)} \ll_\varepsilon t_g^{-\frac{\delta}{2}+\varepsilon}.$$

3.8 Spectral methods to bound the bulk range

In [45, Chapter 6], Spinu proves the bound

$$\sum_{2\alpha t_g \leq t_f \leq 2(1-\alpha)t_g} \frac{\Lambda\left(\frac{1}{2}, f\right)^2 \Lambda\left(\frac{1}{2}+2it_g, f\right) \Lambda\left(\frac{1}{2}-2it_g, f\right)}{\Lambda(1+2it_g)^2 \Lambda(1-2it_g)^2 \Lambda\left(1, \mathrm{sym}^2 f\right)} \ll_\alpha \left(\log\left(\frac{1}{4}+t_g^2\right)\right)^2$$

for any small $\alpha > 0$. Via the methods of Buttcane and Khan [5,6] (the chief difference of which is using a different test function in the Kuznetsov formula), this extends to the full bulk range $t_g^{1-\delta} < t_f < 2t_g - t_g^{1-\delta}$, which thereby completes the unconditional proof of Theorem 1.9.

4 Failure of equidistribution at the Planck scale

4.1 The Selberg–Harish–Chandra transform

For $z, w \in \mathbb{H}$, set

$$u(z,w) := \frac{|z-w|^2}{4\Im(z)\Im(w)} = \sinh^2 \frac{\rho(z,w)}{2},$$

where

$$\rho(z,w) := \log \frac{|z-\overline{w}| + |z-w|}{|z-\overline{w}| - |z-w|}$$

denotes the hyperbolic distance on $\mathbb{H}$. The function $u : \mathbb{H} \times \mathbb{H} \to [0,\infty)$ is a point-pair invariant. From this, a function $k : [0,\infty) \to \mathbb{C}$ gives rise to a point-pair invariant $k(z,w) := k(u(z,w))$ on $\mathbb{H}$. The Selberg–Harish-Chandra transform maps sufficiently well-behaved functions $k : [0,\infty) \to \mathbb{C}$ to functions $h : \mathbb{R} \to \mathbb{C}$. This transform is given in three steps as follows:

$$q(v) := \int_v^\infty \frac{k(u)}{\sqrt{u-v}}\,du, \qquad g(r) := 2q\left(\sinh^2\frac{r}{2}\right), \qquad h(t) := \int_{-\infty}^\infty g(r)e^{irt}\,dr.$$

Note that $h(t)$ is real whenever t is real.

We shall take $k(z, w) = k_R(z, w)$ equal to the indicator function of a small ball of radius R centred at a point w,

$$B_R(w) := \{z \in \mathbb{H} : \rho(z, w) \leq R\} = \left\{z \in \mathbb{H} : u(z, w) \leq \sinh^2\frac{R}{2}\right\},$$

normalised by the volume of this ball. So

$$k(u) = k_R(u) := \begin{cases} \dfrac{1}{4\pi \sinh^2 \frac{R}{2}} & \text{if } u \leq \sinh^2\dfrac{R}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

and consequently

$$h(t) = h_R(t) := \frac{R}{\pi \sinh\frac{R}{2}} \int_{-1}^1 \sqrt{1 - \left(\frac{\sinh\frac{Rr}{2}}{\sinh\frac{R}{2}}\right)^2}\, e^{iRrt}\,dr.$$

We require the following asymptotics for $h_R(t)$, which are extremely similar to the analogous result for $\mathbb{T}^2$; see [15, Lemma 2.1].

Lemma 4.2 (Cf. [7, Lemma 2.4]) *As R tends to zero, we have that*

$$h_R(t) \sim \begin{cases} 1 & \textit{if } Rt \textit{ tends to zero,} \\ \dfrac{2J_1(Rt)}{Rt} & \textit{if } Rt \in (0, \infty), \\ \dfrac{1}{\sqrt{\pi}}\left(\dfrac{2}{Rt}\right)^{3/2} \sin\left(Rt - \dfrac{\pi}{4}\right) & \textit{if } Rt \textit{ tends to infinity.} \end{cases}$$

Proof If R and Rt both converge to zero, then the dominated convergence theorem implies that

$$h_R(t) \sim \frac{2}{\pi}\int_{-1}^1 \sqrt{1-r^2}\,dr = 1.$$

If R converges to 0 and Rt converges to some value in $(0, \infty)$, then similarly

$$h_R(t) \sim \frac{2}{\pi}\int_{-1}^1 \sqrt{1-r^2}e^{iRrt}\,dr = \frac{2J_1(Rt)}{Rt}$$

via [14, 8.411.10]. So it remains to prove the case that R converges to 0 and Rt tends to infinity. To do this, we let

$$h(R,x) := \frac{R}{\pi \sinh \frac{R}{2}} \int_{-1}^{1} \sqrt{1 - \left(\frac{\sinh \frac{Rr}{2}}{\sinh \frac{R}{2}}\right)^2} e^{irx} \, dr.$$

We show that

$$x^{3/2} h(R,x) - 2\sqrt{\frac{2}{\pi} \frac{R}{\sinh R}} \sin\left(x - \frac{\pi}{4}\right)$$

is pointwise convergent as R tends to zero and is uniformly convergent to 0 as x tends to infinity, from which the Moore–Osgood theorem allows us to interchange the order of limits taken in order to obtain the desired asymptotic. Indeed, the dominated convergence theorem once again shows that $h(R,x)$ converges to

$$\frac{2}{\pi} \int_{-1}^{1} \sqrt{1-r^2} e^{irx} \, dr = \frac{2J_1(x)}{x}$$

as R tends to zero. For the uniform convergence as x tends to infinity, we integrate by parts and make the substitution $r = \frac{2}{R} \operatorname{arsinh}\left(\sin v \sinh \frac{R}{2}\right)$, yielding

$$h(R,x) = \frac{R}{2\sinh \frac{R}{2}} \frac{2}{\pi i x} \int_{-\pi/2}^{\pi/2} \sin v e^{ix \frac{2}{R} \operatorname{arsinh}\left(\sin v \sinh \frac{R}{2}\right)} \, dv.$$

Using stationary phase, with the two critical points being the endpoints $\pm\pi/2$, we find that there exists some $R_0 > 0$ such that

$$\sup_{R \in (0,R_0)} \left| x^{3/2} h(R,x) - 2\sqrt{\frac{2}{\pi} \frac{R}{\sinh R}} \sin\left(x - \frac{\pi}{4}\right) \right| \ll \frac{1}{x}.$$

□

For a function $k : [0,\infty) \to \mathbb{C}$, we may form the automorphic kernel

$$K(z,w) := \sum_{\gamma \in \Gamma} k(\gamma z, w),$$

which is Γ-invariant in both variables. When $k(u) = k_R(u)$, we write $K(z,w) = K_R(z,w)$.

Lemma 4.3 *If $f : \Gamma \backslash \mathbb{H} \to \mathbb{C}$ is an eigenfunction of the Laplacian with eigenvalue $1/4 + t_f^2$, then*

$$\frac{1}{\operatorname{vol}(B_R))} \int_{B_R(w)} f(z) \, d\mu(z) = \langle f, K_R(\cdot, w) \rangle = h_R(t_f) f(w).$$

Proof This follows from [28, Theorem 1.14]. Note that there it is assumed that not only is $k(u)$ compactly supported, but that it is smooth; this, however, is not essential to the proof. Instead, we merely require that $k(z, w)$ be twice differentiable in both variables μ-almost everywhere. □

4.2 Proof of theorem 1.14

Proposition 4.4 ([41, Theorem 1]) *For every fixed Heegner point $w \in \mathbb{H}$,*

$$|g(w)| = \Omega\left(\exp\left(\sqrt{\frac{\log t_g}{\log\log t_g}}\left(1 + O\left(\frac{\log\log\log t_g}{\log\log t_g}\right)\right)\right)\right)$$

as t_g tends to infinity.

Proof of Theorem 1.14 For $g \in \mathcal{B}_0(\Gamma)$,

$$\frac{1}{\operatorname{vol}(B_R)} \int_{B_R(w)} g(z)\, d\mu(z) = \int_{\Gamma\backslash\mathbb{H}} K_R(z, w) g(z)\, d\mu(z) = h_R(t_g) g(w).$$

It follows by the Cauchy–Schwarz inequality that

$$\left|h_R\left(t_g\right)\right|^2 |g(w)|^2 \leq \frac{1}{\operatorname{vol}(B_R)} \int_{B_R(w)} |g(z)|^2\, d\mu(z).$$

Theorem 1.14 then follows from Lemma 4.2 and Proposition 4.4. □

Remark 4.5 Theorem 1.14 also holds for Maaß newforms $g \in \mathcal{B}_0^*(\Gamma_0(q))$ for any $q > 1$, for Proposition 4.4 is proved in this generality (and in fact in even further generality).

Remark 4.6 Since it is conjectured that $\max_{w \in K} |g(w)| \ll_{K,\varepsilon} t_g^{\varepsilon}$ for every compact subset K of $\Gamma\backslash\mathbb{H}$, we cannot expect any significant improvement to Theorem 1.14 via this line of reasoning.

5 Equidistribution in almost every shrinking ball

5.1 Proof of conditional results

In this section, we prove the following.

Proposition 5.1 *Let $g \in \mathcal{B}_0(\Gamma)$ be a Hecke–Maaß eigenform normalised such that $\langle g, g\rangle = 1$. For $R > 0$, let*

$$\operatorname{Var}(g; R) := \int_{\Gamma\backslash\mathbb{H}} \left(\frac{1}{\operatorname{vol}(B_R)} \int_{B_R(w)} |g(z)|^2\, d\mu(z) - \frac{1}{\operatorname{vol}(\Gamma\backslash\mathbb{H})}\right)^2 d\mu(w).$$

Assume the generalised Lindelöf hypothesis, and suppose that $R \asymp t_g^{-\delta}$ *for some* $\delta > 0$. *Then for* $0 < \delta < 1$,

$$\mathrm{Var}(g; R) \ll_\varepsilon t_g^{-(1-\delta)+\varepsilon}$$

as t_g *tends to infinity, while for* $\delta > 1$,

$$\mathrm{Var}(g; R) \sim \frac{2}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} = \frac{6}{\pi}$$

as t_g *tends to infinity.*

Theorem 1.17 then follows directly via Chebyshev's inequality. Our starting point towards proving Proposition 5.1 is the following spectral expansion of $\mathrm{Var}(g; R)$.

Proposition 5.2 *Let* $g \in \mathcal{B}_0(\Gamma)$ *be a Hecke–Maaß eigenform normalised such that* $\langle g, g\rangle = 1$. *Then* $\mathrm{Var}(g; R)$ *is equal to*

$$\sum_{f \in \mathcal{B}_0(\Gamma)} \left|h_R(t_f)\right|^2 \left|\left\langle |g|^2, f\right\rangle\right|^2 + \frac{1}{4\pi}\int_{-\infty}^{\infty} |h_R(t)|^2 \left|\left\langle |g|^2, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle\right|^2 dt,$$

where

$$h_R(t) := \frac{R}{\pi \sinh \frac{R}{2}} \int_{-1}^{1} \sqrt{1 - \left(\frac{\sinh \frac{Rr}{2}}{\sinh \frac{R}{2}}\right)^2}\, e^{iRrt}\, dr.$$

Proof Via Lemmata 3.1 (namely Parseval's identity) and 4.3, $\langle |g|^2, K_R(\cdot, w)\rangle$ is equal to

$$\begin{aligned}&\frac{\langle g, g\rangle}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})} + \sum_{f \in \mathcal{B}_0(\Gamma)} h_R(t_f) f(w) \left\langle |g|^2, f\right\rangle \\ &+ \frac{1}{4\pi}\int_{-\infty}^{\infty} h_R(t) E\left(w, \frac{1}{2}+it\right)\left\langle |g|^2, E\left(\cdot, \frac{1}{2}+it\right)\right\rangle dt.\end{aligned}$$

Upon squaring and integrating over w, we obtain the desired identity. □

Proof of Proposition 5.1 *for* $0 < \delta < 1$ We use Propositions 5.2 and 2.8 and Lemmata 4.2 and 2.12. We then divide the spectral expansion in Proposition 5.2 into various ranges.

Just as in Sect. 3.2, there are two main ranges of the continuous spectrum to consider:

- the initial range $0 \leq |t| < 2t_g + t_g^{\delta}$, and
- the tail range $|t| > 2t_g + t_g^{\delta}$.

The division of the cuspidal spectrum into parts depends on δ. When $R \asymp t_g^{-\delta}$ with $0 < \delta < 1$, the ranges are:

- the short initial range $0 < t_f \leq t_g^{\delta}$,
- the polynomial decay range $t_g^{\delta} < t_f < 2t_g + t_g^{1-\delta}$,
- the tail range $t_f \geq 2t_g + t_g^{1-\delta}$.

Thus $\mathrm{Var}(g; R)$ is bounded by a constant multiple dependent on ε of

$$
\begin{aligned}
&t_g^{-1+\varepsilon} \sum_{0<t_f\leq t_g^{\delta}} \frac{L\left(\frac{1}{2}, f\right) L\left(\frac{1}{2}, \mathrm{sym}^2 g \otimes f\right)}{t_f L\left(1, \mathrm{sym}^2 f\right)} \\
&+ t_g^{3\delta-\frac{1}{2}+\varepsilon} \sum_{t_g^{\delta}<t_f<2t_g+t_g^{1-\delta}} \frac{L\left(\frac{1}{2}, f\right) L\left(\frac{1}{2}, \mathrm{sym}^2 g \otimes f\right)}{t_f^4 (1+|2t_g - t_f|)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \\
&+ t_g^{3\delta+\varepsilon} \sum_{t_f\geq 2t_g+t_g^{1-\delta}} e^{-\pi(t_f-2t_g)} \frac{L\left(\frac{1}{2}, f\right) L\left(\frac{1}{2}, \mathrm{sym}^2 g \otimes f\right)}{t_f^{\frac{9}{2}} (1+t_f - 2t_g)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \\
&+ t_g^{-\frac{1}{2}+\varepsilon} \int_0^{2t_g+t_g^{\delta}} \frac{\left|L\left(\frac{1}{2}+it\right) L\left(\frac{1}{2}+it, \mathrm{sym}^2 g\right)\right|^2}{(1+t)(1+|2t_g - t|)^{1/2} |\zeta(1+2it)|^2}\, dt \\
&+ t_g^{-\frac{1}{2}+\varepsilon} \int_{2t_g+t_g^{\delta}}^{\infty} e^{-\pi(t-2t_g)} \frac{\left|L\left(\frac{1}{2}+it\right) L\left(\frac{1}{2}+it, \mathrm{sym}^2 g\right)\right|^2}{(1+t)(1+|2t_g - t|)^{1/2} |\zeta(1+2it)|^2}\, dt.
\end{aligned}
$$

- From [6, Lemma 2.1], the initial and tail ranges of the continuous spectrum are bounded by $t_g^{-1+\varepsilon}$.
- The convexity bounds for $L(1/2, f)$ and $L(1/2, \mathrm{sym}^2 g \otimes f)$ show that the tail range of the cuspidal spectrum is rapidly decaying.
- For the other two ranges, the generalised Lindelöf hypothesis implies that the product of these two L-functions is bounded by a constant multiple dependent on ε of t_g^{ε}, and then the Weyl law for $\Gamma\backslash\mathbb{H}$ and partial summation imply that the contribution of the cuspidal spectrum is bounded by $t_g^{\delta-1+\varepsilon}$.

This completes the proof. □

Proof of Proposition 5.1 *for* $\delta > 1$ In this case, the division of the cuspidal spectrum into parts involves an additional range, and there is a dependence on an small fixed parameter $\delta' > 0$:

- the short initial range $0 < t_f \leq t_g^{1-\delta'}$, which once again is bounded by $t_g^{-\delta'/2+\varepsilon}$ via the generalised Lindelöf hypothesis,
- the bulk range $t_g^{1-\delta'} < t_f < 2t_g - t_g^{1-\delta'}$, which is asymptotic to $6/\pi$ from the proof of [6, Proposition 2.2],
- the short transition range $2t_g - t_g^{1-\delta'} \leq t_f \leq 2t_g + t_g^{1-\delta'}$, again bounded by $t_g^{-\delta'/2+\varepsilon}$, and
- the tail range $t_f > 2t_g + t_g^{1-\delta'}$, which is negligible.

This completes the proof. □

Remark 5.3 Just as with Theorem 1.14, the bound $\mathrm{Var}(g; R) \ll_\varepsilon t_g^{-(1-\delta)+\varepsilon}$ for $R \asymp t_g^{-\delta}$ with $0 < \delta < 1$ in Proposition 5.1 also holds for Maaß newforms $g \in \mathcal{B}_0^*(\Gamma_0(q))$ for any $q > 1$. Indeed, [29, Theorem 15.5] gives the spectral decomposition of $L^2(\Gamma_0(q)\backslash\mathbb{H})$, though there are Eisenstein series corresponding to each cusp and the orthonormal basis of Maaß cusp forms are no longer necessarily Hecke–Maaß eigenforms. Nonetheless, Blomer and Milićević have given an orthonormal basis of $\mathcal{B}_0(\Gamma_0(q))$ involving linear combinations of oldforms and newforms [2, Lemma 9], and a similar basis exists for the space of Eisenstein series [49], and these can be coupled with the work of Hu on the Watson–Ichino formula in this generality [25].

Remark 5.4 In fact, the method of proof of [6, Proposition 2.2] together with Lemma 4.2 show that if $R \sim (Ct_g)^{-1}$ for some positive constant C, then

$$\mathrm{Var}(g; R) \sim \frac{12C}{\pi^2} \int_0^1 \frac{J_1\left(\frac{2t}{C}\right)}{t\sqrt{1-t^2}}\, dt = \frac{6}{\pi}\left(J_0\left(\frac{1}{C}\right)^2 + J_1\left(\frac{1}{C}\right)^2\right)$$

by [14, (8.473.1) and (6.552.4)], which converges to $6/\pi$ as C tends to infinity.

5.2 Proof of unconditional results

We first sketch how to prove Theorem 1.16.

Proof of Theorem 1.16 In [47], after [47, (4.24)], we use Lemma 3.10 instead of the subconvexity bound $L(1/2+it, f) \ll_\varepsilon (t_f + t)^{1/3+\varepsilon}$. Using this, the right-hand side of [47, (4.26)] is improved to $T^{-1/6+\varepsilon}\|\phi\|_2$, which yields the result. □

Next, we cover the proof of the following, from which Theorem 1.17 will be derived.

Proposition 5.5 *Let $g(z) = E\left(z, 1/2 + it_g\right)$. For $R > 0$, let*

$$\mathrm{Var}(g; R) := \int_{\Gamma\backslash\mathbb{H}} \left(\frac{1}{\mathrm{vol}\,(B_R)} \int_{B_R(w)} |g(z)|^2\, d\mu(z) - C(g; R; w)\right)^2 d\mu(w),$$

where $C(g; R; w)$ is given by (5.8). Suppose that $R \asymp t_g^{-\delta}$ for some $0 < \delta < 1$. Then

$$\mathrm{Var}(g; R) \ll_\varepsilon t_g^{-\min\left\{\frac{5}{7}(1-\delta), \frac{1}{6}\right\}+\varepsilon}.$$

To begin, we wish to calculate

$$\frac{1}{\mathrm{vol}(B_R)} \int_{B_R(w)} |g(z)|^2\, d\mu(z),$$

where $g(z) = E\left(z, 1/2 + it_g\right)$. However, we cannot use Parseval's identity because $|g|^2 \notin L^2\left(\Gamma\backslash\mathbb{H}\right)$. Instead, we replace $|g(z)|^2$ with $E(z, s_1)E(z, s_2)$ and subtract away a linear combination of Eisenstein series $\mathcal{E}$ such that the resulting function is square-integrable. After applying Parseval's identity, we finally send s_1 to $1/2+it_g$ and s_2 to $1/2-it_g$.

Lemma 5.6 (Cf. [47, Lemma 4.1]) *For* $1/2<\Re(s_1), \Re(s_2)<3/4$,

$$\frac{1}{\mathrm{vol}(B_R)}\int_{B_R(w)} E(z,s_1)E(z,s_2)\,d\mu(z)$$

is equal to

$$\begin{aligned}
&h_R\left(i\left(s_1+s_2-\frac{1}{2}\right)\right)E(w,s_1+s_2)\\
&\quad+h_R\left(i\left(\frac{1}{2}-s_1+s_2\right)\right)\frac{\Lambda(2-2s_1)}{\Lambda(2s_1)}E(w,1-s_1+s_2)\\
&\quad+h_R\left(i\left(\frac{1}{2}+s_1-s_2\right)\right)\frac{\Lambda(2-2s_2)}{\Lambda(2s_2)}E(w,1+s_1-s_2)\\
&\quad+h_R\left(i\left(\frac{3}{2}-s_1-s_2\right)\right)\frac{\Lambda(2-2s_1)\Lambda(2-2s_2)}{\Lambda(2s_1)\Lambda(2s_2)}E(w,2-s_1-s_2)\\
&\quad+\sum_{f\in\mathcal{B}_0(\Gamma)} h_R(t_f)f(w)\,\langle E(\cdot,s_1)E(\cdot,s_2),f\rangle\\
&\quad+\frac{1}{4\pi}\int_{-\infty}^{\infty}h_R(t)E\left(w,\frac{1}{2}+it\right)\left\langle E(\cdot,s_1)E(\cdot,s_2),E\left(\cdot,\frac{1}{2}+it\right)\right\rangle_{\mathrm{reg}}dt.
\end{aligned}$$

Proof Let $F(z):=E(z,s_1)E(z,s_2)$ and let

$$\begin{aligned}
\mathcal{E}(z):=E(z,s_1+s_2)+\frac{\Lambda(2-2s_1)}{\Lambda(2s_1)}E(z,1-s_1+s_2)\\
+\frac{\Lambda(2-2s_2)}{\Lambda(2s_2)}E(z,1+s_1-s_2)+\frac{\Lambda(2-2s_1)\Lambda(2-2s_2)}{\Lambda(2s_1)\Lambda(2s_2)}E(z,2-s_1-s_2).
\end{aligned}$$

Since the constant term of $F(z)$ is

$$y^{s_1+s_2}+\frac{\Lambda(2-2s_1)}{\Lambda(2s_1)}y^{1-s_1+s_2}+\frac{\Lambda(2-2s_2)}{\Lambda(2s_2)}y^{1+s_1-s_2}+\frac{\Lambda(2-2s_1)\Lambda(2-2s_2)}{\Lambda(2s_1)\Lambda(2s_2)}y^{2-s_1-s_2},$$

we have that $F(z)-\mathcal{E}(z)=O(y^{1/2-\delta})$ for some $\delta>0$ at the cusp at infinity, and consequently $F-\mathcal{E}\in L^2(\Gamma\backslash\mathbb{H})$. Lemmata 3.1 (namely Parseval's identity) and 4.3 then imply that

$$\begin{aligned}
\langle F-\mathcal{E},K_R(\cdot,w)\rangle=\frac{\langle F-\mathcal{E},1\rangle}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}+\sum_{f\in\mathcal{B}_0(\Gamma)}h_R(t_f)f(w)\langle F-\mathcal{E},f\rangle\\
+\frac{1}{4\pi}\int_{-\infty}^{\infty}h_R(t)E\left(w,\frac{1}{2}+it\right)\left\langle F-\mathcal{E},E\left(\cdot,\frac{1}{2}+it\right)\right\rangle dt.
\end{aligned}$$

The left-hand side is equal to $\langle F, K_R(\cdot, w)\rangle - \langle \mathcal{E}, K_R(\cdot, w)\rangle$, and Lemma 4.3 allows us to calculate $\langle \mathcal{E}, K_R(\cdot, w)\rangle$ explicitly. On the right-hand side, the inner product $\langle \mathcal{E}, f\rangle$ vanishes whenever $f \in \mathcal{B}_0(\Gamma)$, being the linear combination of inner products of Eisenstein series with a cusp form, and similarly $\langle F - \mathcal{E}, 1\rangle$ vanishes via [50, Equation (36) and Sect. 2]. Finally, we claim that the inner product $\langle F - \mathcal{E}, E\left(\cdot, \frac{1}{2} + it\right)\rangle$ is equal to

$$\frac{\Lambda\left(s_2 - s_1 + \frac{1}{2} + it\right)\Lambda\left(s_1 + s_2 - \frac{1}{2} + it\right)\Lambda\left(s_2 - s_1 + \frac{1}{2} - it\right)\Lambda\left(s_1 + s_2 - \frac{1}{2} - it\right)}{\Lambda\left(2s_1\right)\Lambda\left(2s_2\right)\Lambda\left(1 - 2it\right)}.$$

Indeed, we may add and subtract a linear combination of Eisenstein series $\mathcal{E}'$ such that both $FE\left(\cdot, 1/2 - it\right) - \mathcal{E}'$ and $\mathcal{E}E\left(\cdot, 1/2 - it\right) - \mathcal{E}'$ are integrable. Then the integral of $\mathcal{E}E\left(\cdot, 1/2 + it\right) - \mathcal{E}'$ vanishes via [50, Equation (36) and Sect. 2], and the integral of $FE\left(\cdot, 1/2 + it\right) - \mathcal{E}'$ is equal to the desired product of completed zeta functions via [50, Equation (44)]. $\square$

We now define

$$D(g; w) := \frac{2}{\operatorname{vol}\left(\Gamma\backslash\mathbb{H}\right)}\left(2\Re\left(\frac{\Lambda'}{\Lambda}\left(1 + 2it_g\right)\right) + 2\gamma_0 - \frac{12\zeta'(2)}{\pi^2} - \log\left|4\Im(w)\eta(w)^4\right|\right). \tag{5.7}$$

Here γ_0 is the Euler–Mascheroni constant and

$$\eta(w) := e\left(\frac{w}{24}\right)\prod_{m=1}^{\infty}\left(1 - e(mw)\right)$$

denotes the Dedekind eta function; note that $\Im(w)^6\eta(w)^{24}$ is a Maaß cusp form of weight 12 and level 1 that is nonvanishing outside the single cusp of $\Gamma\backslash\mathbb{H}$. That $D(g; w)$ is, in some sense, the "true" average of $|E(z, 1/2 + it_g)|^2$ on compact sets, rather than

$$\frac{\log\left(\frac{1}{4} + t_g^2\right)}{\operatorname{vol}\left(\Gamma\backslash\mathbb{H}\right)},$$

has previously been observed by Young [47, Sect. 4.2] and also Hejhal and Rackner [20, p. 300], though in the latter case, their expression does not include the Dedekind eta function.

Proof of Lemma 1.19 This follows from (2.4), (2.5), and (2.6), together with the fact that $\Im(w)^6\eta(w)^{24}$ is nonvanishing in K. $\square$

We define

$$C(g; R; w) := D(g; w) + \frac{2ih_R'\left(\frac{i}{2}\right)}{\operatorname{vol}\left(\Gamma\backslash\mathbb{H}\right)} + 2\Re\left(h_R\left(2t_g + \frac{i}{2}\right)\frac{\Lambda(1 - 2it_g)}{\Lambda(1 + 2it_g)}E\left(w, 1 - 2it_g\right)\right). \tag{5.8}$$

Lemma 5.9 *Let* $g(z) = E\left(z, 1/2 + it_g\right)$. *Then*

$$\frac{1}{\mathrm{vol}(B_R)} \int_{B_R(w)} |g(z)|^2 \, d\mu(z) = C(g; R; w) + \sum_{f \in \mathcal{B}_0(\Gamma)} h_R(t_f) f(w) \left\langle |g|^2, f \right\rangle$$
$$+ \frac{1}{4\pi} \int_{-\infty}^{\infty} h_R(t) E\left(w, \frac{1}{2} + it\right) \left\langle |g|^2, E\left(\cdot, \frac{1}{2} + it\right)\right\rangle_{\mathrm{reg}} dt.$$

Proof This follows from Lemma 5.6 upon taking $s_1 = 1/2 + it_g + \varepsilon$ and $s_2 = 1/2 - it_g + \varepsilon$ and using the expansions

$$h_R\left(i\left(\frac{1}{2} + 2\varepsilon\right)\right) = 1 + 2ih'_R\left(\frac{i}{2}\right)\varepsilon + O(\varepsilon^2),$$
$$\frac{\Lambda(1 - 2it_g - 2\varepsilon)\Lambda(1 + 2it_g - 2\varepsilon)}{\Lambda(1 + 2it_g + 2\varepsilon)\Lambda(1 - 2it_g + 2\varepsilon)} = 1 - 8\Re\left(\frac{\Lambda'}{\Lambda}\left(1 + 2it_g\right)\right)\varepsilon + O(\varepsilon^2),$$
$$\mathrm{vol}\,(\Gamma\backslash\mathbb{H})\, E(w, 1 + 2\varepsilon) = \frac{1}{2\varepsilon} + 2\gamma_0 - \log\left|4\Im(w)\eta(w)^4\right| - \frac{12\zeta'(2)}{\pi^2} + O(\varepsilon),$$

where the last line is the Kronecker limit formula. □

With this in hand, we can finally give the spectral expansion of $\mathrm{Var}(g; R)$.

Proposition 5.10 *Let* $g(z) = E\left(z, 1/2 + it_g\right)$. *Then* $\mathrm{Var}(g; R)$ *is equal to*

$$\sum_{f \in \mathcal{B}_0(\Gamma)} \left|h_R(t_f)\right|^2 \left|\left\langle |g|^2, f\right\rangle\right|^2 + \frac{1}{4\pi} \int_{-\infty}^{\infty} |h_R(t)|^2 \left|\left\langle |g|^2, E\left(\cdot, \frac{1}{2} + it\right)\right\rangle_{\mathrm{reg}}\right|^2 dt.$$

Proof This follows directly from Lemma 5.9 after an application of Parseval's identity in Lemma 3.1. □

Proof of Proposition 5.5 We use Propositions 5.10 and 2.9 and Lemmata 4.2 and 2.12. We then divide the spectral expansion in Proposition 5.10 into various ranges.

The two ranges of the continuous spectrum are:

- the initial range $0 \le |t| < 2t_g + t_g^{\delta}$, and
- the tail range $|t| > 2t_g + t_g^{\delta}$.

The cuspidal spectrum can be broken into five ranges, which depend on a small fixed parameter $0 < \delta' < 1 - \delta$:

- the short initial range $0 < t_f \le t_g^{\delta}$,
- the short initial polynomial decay range $t_g^{\delta} < t_f < t_g^{1-\delta'}$,
- the bulk polynomial decay range $t_g^{1-\delta'} \le t_f \le 2t_g - t_g^{\delta}$,
- the short transition polynomial decay range $2t_g - t_g^{\delta} < t_f < 2t_g + t_g^{\delta}$,

- the tail range $t_f \geq 2t_g + t_g^{\delta}$.

Thus $\mathrm{Var}(g; R)$ is bounded by a constant multiple dependent on ε of

$$
\begin{aligned}
& t_g^{-1+\varepsilon} \sum_{0<t_f\leq t_g^{\delta}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{t_f L\left(1, \mathrm{sym}^2 f\right)} \\
& + t_g^{3\delta-1+\varepsilon} \sum_{t_g^{\delta}<t_f<t_g^{1-\delta'}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{t_f^4 L\left(1, \mathrm{sym}^2 f\right)} \\
& + t_g^{3\delta-\frac{1}{2}+\varepsilon} \sum_{t_g^{1-\delta'}\leq t_f\leq 2t_g-t_g^{\delta}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{t_f^4 (1+|2t_g-t_f|)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \\
& + t_g^{3\delta-\frac{9}{2}+\varepsilon} \sum_{2t_g-t_g^{\delta}<t_f<2t_g+t_g^{\delta}} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{(1+|2t_g-t_f|)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \\
& + t_g^{3\delta+\varepsilon} \sum_{t_f\geq 2t_g+t_g^{\delta}} e^{-\pi(t_f-2t_g)} \frac{L\left(\frac{1}{2}, f\right)^2 \left|L\left(\frac{1}{2}+2it_g, f\right)\right|^2}{t_f^{\frac{9}{2}} (1+t_f-2t_g)^{1/2} L\left(1, \mathrm{sym}^2 f\right)} \\
& + t_g^{-\frac{1}{2}+\varepsilon} \int_0^{2t_g+t_g^{\delta}} \frac{\left|\zeta\left(\frac{1}{2}+i(2t_g+t)\right) \zeta\left(\frac{1}{2}+it\right)^2 \zeta\left(\frac{1}{2}+i(2t_g-t)\right)\right|^2}{(1+t)(1+|2t_g-t|)^{1/2} \left|\zeta(1-2it)\right|^2}\, dt \\
& + t_g^{-\frac{1}{2}+\varepsilon} \int_{2t_g+t_g^{\delta}}^{\infty} e^{-\pi(t-2t_g)} \frac{\left|\zeta\left(\frac{1}{2}+i(2t_g+t)\right) \zeta\left(\frac{1}{2}+it\right)^2 \zeta\left(\frac{1}{2}+i(2t_g-t)\right)\right|^2}{(1+t)(1+|2t_g-t|)^{1/2} \left|\zeta(1-2it)\right|^2}\, dt.
\end{aligned}
$$

The continuous spectrum is readily dealt with:

- From [45, Proposition 3.4] and [3, Theorem 5], the initial and tail ranges of the continuous spectrum are bounded by a constant multiple dependent on ε of $t_g^{-\frac{13}{84}+\varepsilon}$.

For the cuspidal spectrum, we have the following:

- The convexity bounds for $L(1/2, f)$ and $L(1/2+2it_g, f)$ show that the tail range is rapidly decaying.
- The short initial range is bounded by a constant multiple dependent on ε of $t_g^{-\min\{1-\delta,1/6\}+\varepsilon}$ upon dividing into dyadic intervals and applying Lemma 3.10.
- The same method bounds the short initial polynomial decay range by $t_g^{-\min\{\delta',1/6\}+\varepsilon}$.
- For the bulk polynomial decay range, we divide into dyadic intervals and use Lemma 3.7, which shows that this range is bounded by $t_g^{-\frac{5}{2}(1-\delta-\delta')+\varepsilon}$.
- We divide the short transition polynomial decay range into intervals of length $t_g^{1/3}$, use the Cauchy–Schwarz inequality, and apply Lemma 3.12, which gives the bound $t_g^{-\frac{7}{2}(1-\delta)+\varepsilon}$.

Proposition 5.5 is proven upon taking $\delta' = \frac{5}{7}(1-\delta)$. □

Proof of Theorem 1.18 By Chebyshev's inequality and Proposition 5.5,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{1}{\mathrm{vol}(B_R)}\int_{B_R(w)} |g(z)|^2 \, d\mu(z) - C(g; R; w)\right| > c\right\}\right)$$
$$\ll_\varepsilon \frac{t_g^{-\min\left\{\frac{5}{7}(1-\delta), \frac{1}{6}\right\}+\varepsilon}}{c^2}.$$

Again by Chebyshev's inequality,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|h_R\left(2t_g + \frac{i}{2}\right) E\left(w, 1 - 2it_g\right)\right| > c\right\}\right)$$
$$\leq \mathrm{vol}\left(\{w \in \Gamma\backslash\mathbb{H} : \Im(w) > T\}\right) + \frac{\left|h_R\left(2t_g + \frac{i}{2}\right)\right|^2}{c^2} \int_{\Gamma\backslash\mathbb{H}} \left|\Lambda^T E\left(w, 1 - 2it_g\right)\right|^2 d\mu(z)$$

for any $T \geq 1$, which, by the Maaß–Selberg relation (2.2), is equal to

$$\frac{1}{T} + \frac{\left|h_R\left(2t_g + \frac{i}{2}\right)\right|^2}{c^2}\left(T + 2\Re\left(\frac{\Lambda(1-4it_g)}{\Lambda(2-4it_g)}\frac{T^{4it_g}}{4it_g}\right) + \left|\frac{\Lambda(1-4it_g)}{\Lambda(2-4it_g)}\right|^2 \frac{1}{T}\right).$$

Using stationary phase as in the proof of Lemma 4.2, or alternatively using [7, Lemma 2.4], we have that $|h_R\left(2t_g + \frac{i}{2}\right)|^2 \ll t_g^{-3(1-\delta)}$, while Stirling's approximation implies that

$$\frac{\Lambda(1-2it_g)}{\Lambda(2-4it_g)} \ll_\varepsilon t_g^{-\frac{1}{2}+\varepsilon}.$$

Next, we note that

$$ih_R'\left(\frac{i}{2}\right) = \frac{R^2}{\pi}\int_{-1}^{1} r\sqrt{1 - \left(\frac{\sinh\frac{Rr}{2}}{\sinh\frac{R}{2}}\right)^2}\frac{\sinh\frac{Rr}{2}}{\sinh\frac{R}{2}}\, dr \sim \frac{R^2}{8} \asymp t_g^{-2\delta},$$

so if $c \gg_\varepsilon t_g^{-2\delta+\varepsilon}$, then for all sufficiently large t_g,

$$\left|\frac{2ih_R'\left(\frac{i}{2}\right)}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right| < c.$$

So piecing everything together, we find that if $c \gg_\varepsilon t_g^{-2\delta+\varepsilon}$,

$$\mathrm{vol}\left(\left\{w \in \Gamma\backslash\mathbb{H} : \left|\frac{1}{\mathrm{vol}(B_R)}\int_{B_R(w)} |g(z)|^2 \, d\mu(z) - D(g; w)\right| > c\right\}\right)$$
$$\ll_\varepsilon \frac{t_g^{-\frac{5}{7}(1-\delta)+\varepsilon}}{c^2} + \frac{t_g^{-\frac{1}{6}+\varepsilon}}{c^2} + \frac{1}{T} + \frac{t_g^{-3(1-\delta)}T}{c^2}.$$

Taking $T = ct_g^{\frac{3}{2}(1-\delta)}$ yields the result. $\square$

6 Equidistribution of geometric invariants of quadratic fields

6.1 Geometric invariants of quadratic fields

Let $K = \mathbb{Q}(\sqrt{D})$ be a quadratic field of discriminant D. We denote by $h_K^+ := \#\,\mathrm{Cl}_K^+$ the narrow class number of K and $h_K := \#\,\mathrm{Cl}_K$ the (wide) class number of K; note that $\mathrm{Cl}_K^+ = \mathrm{Cl}_K$, so that $h_K^+ = h_K$, except when $D > 1$ and $\mathcal{O}_K^\times$ contains no elements of norm -1, in which case $h_K^+ = 2h_K$. Each narrow ideal class A of Cl_K^+ is associated to an $\mathrm{SL}_2(\mathbb{Z})$-equivalence class of binary quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ of discriminant D.

Associated to equivalence classes of binary quadratic forms are geometric invariants: if $D < 0$, this is a Heegner point $z_A \in \Gamma\backslash\mathbb{H}$, while if $D > 0$, these are a closed geodesic $\mathcal{C}_A \subset \Gamma\backslash\mathbb{H}$ and a hyperbolic orbifold $\Gamma_A\backslash\mathcal{N}_A$ whose boundary is $\mathcal{C}_A$. This last geometric invariant was introduced by Duke, Imamoḡlu, and Tóth in [11].

6.1.1 *Heegner points z_A*

Given a binary quadratic form $Q(x, y) = ax^2 + bxy + cy^2$ of discriminant $b^2 - 4ac = D < 0$, the point

$$z = \frac{-b + i\sqrt{-D}}{2a}$$

lies in $\mathbb{H}$. The equivalence class of binary quadratic forms containing $Q(x, y)$, and hence the corresponding ideal class $A \in \mathrm{Cl}_K$, thereby corresponds to a point $z = z_A$ in $\Gamma\backslash\mathbb{H}$, which we call a Heegner point.

6.1.2 *Closed geodesics $\mathcal{C}_A$*

Given a binary quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ of discriminant $b^2 - 4ac = D > 0$, the points

$$\frac{-b \pm \sqrt{D}}{2a}$$

determine the endpoints of a closed geodesic in $\mathbb{H}$. The equivalence class of binary quadratic forms containing $Q(x, y)$ thereby corresponds to a closed geodesic $\mathcal{C} = \mathcal{C}_A$ in $\Gamma\backslash\mathbb{H}$. The length

$$\ell(\mathcal{C}_A) := \int_{\mathcal{C}_A} ds$$

of $\mathcal{C}_A$, with $ds^2 = y^{-2}dx^2 + y^{-2}dy^2$, is equal to $2\log \epsilon_K^+$, where $\epsilon_K^+ > 1$ is the smallest unit with positive norm in the ring of integers $\mathcal{O}_K$ of K, so that $\epsilon_K^+ = (x + y\sqrt{D})/2$ with $(x, y) \in \mathbb{R}_+^2$ the fundamental solution to the Pell equation $x^2 - Dy^2 = 4$. Note that ϵ_K^+ is equal to ϵ_K, the fundamental unit of K, if $\mathcal{O}_K^\times$ contains no elements of norm -1, whereas $\epsilon_K^+ = \epsilon_K^2$ if $\mathcal{O}_K^\times$ does contain elements of norm -1.

6.1.3 Hyperbolic orbifolds $\Gamma_A\backslash\mathcal{N}_A$

Let $K = \mathbb{Q}(\sqrt{D})$ be a real quadratic field of discriminant $D > 1$. Associated to a narrow ideal class $A \in \mathrm{Cl}_K^+$ is an invariant $((n_1, \ldots, n_{\ell_A}))$, where ℓ_A is a positive integer and $n_1, \ldots, n_{\ell_A}$ are integers; this is the primitive cycle, unique up to cyclic permutations, occurring in the minus continued fraction expansion of each point $w \in K$ for which $1 > w > \sigma(w) > 0$ and $w\mathbb{Z} + \mathbb{Z} \in A$. We define the elements

$$S := \pm\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad T := \pm\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

of $\mathrm{PSL}_2(\mathbb{Z})$, which generate $\mathrm{PSL}_2(\mathbb{Z})$ as the free product of S and T. For each $k \in \{1, \ldots, \ell_A\}$, define

$$S_k := T^{n_1+\cdots+n_k} S T^{-n_1-\cdots-n_k}.$$

This is an elliptic element of order 2 in $\mathrm{PSL}_2(\mathbb{Z})$. We set

$$\Gamma_A := \left\langle S_1, \cdots, S_{\ell_A}, T^{n_1+\cdots+n_{\ell_A}} \right\rangle,$$

which is a thin subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. The Nielsen region $\mathcal{N}_A$ of Γ_A is the smallest nonempty $\mathrm{PSL}_2(\mathbb{Z})$-invariant open convex subset of $\mathbb{H}$. Then $\Gamma_A\backslash\mathcal{N}_A$ is a hyperbolic orbifold, which naturally projects onto $\Gamma\backslash\mathbb{H}$. The boundary of $\Gamma_A\backslash\mathcal{N}_A$ is a simple closed geodesic whose image in $\Gamma\backslash\mathbb{H}$ is $\mathcal{C}_A$, and the volume of $\Gamma_A\backslash\mathcal{N}_A$ is $\pi\ell_A$.

Remark 5.1 In fact, Γ_A depends on the choice of w. The resulting hyperbolic orbifold $\Gamma_A\backslash\mathcal{N}_A$ ends up being only unique up to translation; however, the projection of $\Gamma_A\backslash\mathcal{N}_A$ onto $\Gamma\backslash\mathbb{H}$ is independent of the choice of w.

6.2 Weyl sums

6.2.1 Variances and Weyl sums

We define

$$\mathrm{Var}\,(G_K(z_A); R) := \int_{\Gamma\backslash\mathbb{H}} \left(\frac{\#\{A \in G_K : z_A \in B_R(w)\}}{\mathrm{vol}\,(B_R)\,\#G_K} - \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}\right)^2 d\mu(w),$$

$$\mathrm{Var}\,(G_K(\mathcal{C}_A); R) := \int_{\Gamma\backslash\mathbb{H}} \left(\frac{\sum_{A\in G_K} \ell\,(\mathcal{C}_A \cap B_R(w))}{\mathrm{vol}\,(B_R)\sum_{A\in G_K} \ell\,(\mathcal{C}_A)} - \frac{1}{\mathrm{vol}\,(\Gamma\backslash\mathbb{H})}\right)^2 d\mu(w),$$

$$\begin{aligned}&\mathrm{Var}\left(G_K(\Gamma_A\backslash\mathcal{N}_A);\,R\right)\\&:=\int_{\Gamma\backslash\mathbb{H}}\left(\frac{\sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\cap B_R(w)\right)}{\mathrm{vol}\left(B_R\right)\sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right)}-\frac{1}{\mathrm{vol}\left(\Gamma\backslash\mathbb{H}\right)}\right)^2\,d\mu(w).\end{aligned}$$

The proofs of Theorems 1.25 and 1.26 follow via Chebyshev's inequality from the following two propositions.

Proposition 5.2 *Suppose that $R \asymp |D|^{-\delta}$. Assuming the generalised Lindelöf hypothesis, we have that as $D \to -\infty$ along fundamental discriminants,*

$$\mathrm{Var}\left(G_K(z_A);\,R\right) \ll_\varepsilon (-D)^{-\left(\frac{1}{4}-\delta\right)+\varepsilon} \quad \text{for } 0 < \delta < 1/4,$$

while as $D \to \infty$ along fundamental discriminants,

$$\mathrm{Var}\left(G_K(\mathcal{C}_A);\,R\right) \ll_\varepsilon D^{-\left(\frac{1}{2}-\delta\right)+\varepsilon} \quad \text{for } 0 < \delta < 1/2.$$

Proposition 5.3 *Suppose that $R \asymp |D|^{-\delta}$. Then as $D \to -\infty$ along odd fundamental discriminants,*

$$\mathrm{Var}\left(G_K(z_A);\,R\right) \ll_\varepsilon (-D)^{-\left(\frac{1}{12}-\delta\right)+\varepsilon} \quad \text{for } 0 < \delta < 1/12,$$

while as $D \to \infty$ along odd fundamental discriminants,

$$\begin{aligned}\mathrm{Var}\left(G_K(\mathcal{C}_A);\,R\right) &\ll_\varepsilon D^{-\left(\frac{1}{6}-\delta\right)+\varepsilon} \quad \text{for } 0 < \delta < 1/6,\\ \mathrm{Var}\left(G_K(\Gamma_A\backslash\mathcal{N}_A);\,R\right) &\ll_\varepsilon D^{-\frac{1}{2}+\varepsilon} \quad \text{for all } \delta > 0.\end{aligned}$$

We begin by determining the spectral expansions of these variances. For $f \in \mathcal{B}_0(\Gamma)$, we define the Weyl sums

$$\begin{aligned}W_{G_K(z_A),f} &:= \sum_{A\in G_K} f\left(z_A\right),\\ W_{G_K(\mathcal{C}_A),f} &:= \sum_{A\in G_K}\int_{\mathcal{C}_A} f(z)\,ds,\\ W_{G_K(\Gamma_A\backslash\mathcal{N}_A),f} &:= \sum_{A\in G_K}\int_{\Gamma_A\backslash\mathcal{N}_A} f(z)\,d\mu(z).\end{aligned}$$

We define $W_{G_K(z_A),\infty}(t)$, $W_{G_K(\mathcal{C}_A),\infty}(t)$, and $W_{G_K(\Gamma_A\backslash\mathcal{N}_A),\infty}(t)$ similarly with f replaced by $E(\cdot, 1/2+it)$.

Proposition 5.4 *We have that*

$$\begin{aligned}
\mathrm{Var}\left(G_K(z_A);\, R\right) &= \sum_{f\in\mathcal{B}_0(\Gamma)} \left|h_R\left(t_f\right)\right|^2 \frac{\left|W_{G_K(z_A),f}\right|^2}{(\#G_K)^2} \\
&\quad + \frac{1}{4\pi}\int_{-\infty}^{\infty} |h_R(t)|^2 \frac{\left|W_{G_K(z_A),\infty}(t)\right|^2}{(\#G_K)^2}\, dt, \\
\mathrm{Var}\left(G_K(\mathcal{C}_A);\, R\right) &= \sum_{f\in\mathcal{B}_0(\Gamma)} \left|h_R\left(t_f\right)\right|^2 \frac{\left|W_{G_K(\mathcal{C}_A),f}\right|^2}{\left(\sum_{A\in G_K}\ell\left(\mathcal{C}_A\right)\right)^2} \\
&\quad + \frac{1}{4\pi}\int_{-\infty}^{\infty} |h_R(t)|^2 \frac{\left|W_{G_K(\mathcal{C}_A),\infty}(t)\right|^2}{\left(\sum_{A\in G_K}\ell\left(\mathcal{C}_A\right)\right)^2}\, dt, \\
\mathrm{Var}\left(G_K(\Gamma_A\backslash\mathcal{N}_A);\, R\right) &= \sum_{f\in\mathcal{B}_0(\Gamma)} \left|h_R\left(t_f\right)\right|^2 \frac{\left|W_{G_K(\Gamma_A\backslash\mathcal{N}_A),f}\right|^2}{\left(\sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right)\right)^2} \\
&\quad + \frac{1}{4\pi}\int_{-\infty}^{\infty} |h_R(t)|^2 \frac{\left|W_{G_K(\Gamma_A\backslash\mathcal{N}_A),\infty}(t)\right|^2}{\left(\sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right)\right)^2}\, dt.
\end{aligned}$$

Proof This follows from the spectral expansion of K_R and Parseval's identity. □

To bound these variances, we require upper bounds for the Weyl sums as well as lower bounds for $\#G_K$, $\sum_{A\in G_K}\ell\left(\mathcal{C}_A\right)$, and $\sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right)$.

Lemma 5.5 *We have that*

$$\begin{aligned}
(-D)^{\frac{1}{2}-\varepsilon} &\ll_\varepsilon \#G_K \ll \sqrt{-D}\log(-D), \\
D^{\frac{1}{2}-\varepsilon} &\ll_\varepsilon \sum_{A\in G_K}\ell\left(\mathcal{C}_A\right) \ll \sqrt{D}\log D, \\
D^{\frac{1}{2}-\varepsilon} &\ll_\varepsilon \sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right) \ll \sqrt{D}\log D.
\end{aligned}$$

Proof We have that $\#G_K = 2^{1-\omega(|D|)}h_K^+$ and $\ell(\mathcal{C}_A) = 2\log\epsilon_K^+$, while it is shown in [11, Proposition 1] that

$$\frac{\#G_K\log\epsilon_K^+}{\log D} \ll \sum_{A\in G_K}\mathrm{vol}\left(\Gamma_A\backslash\mathcal{N}_A\right) \ll \#G_K\log\epsilon_K^+.$$

The class number formula states that

$$h_K^+ = \begin{cases} \dfrac{\sqrt{D}L\left(1,\chi_D\right)}{\log\epsilon_K^+} & \text{if } D > 0, \\[2ex] \dfrac{w_K\sqrt{-D}L\left(1,\chi_D\right)}{2\pi} & \text{if } D < 0, \end{cases}$$

where

$$w_K := \#\mathcal{O}_{K,\mathrm{tors}}^{\times} = \begin{cases} 4 & \text{if } D = -4, \\ 6 & \text{if } D = -3, \\ 2 & \text{otherwise.} \end{cases}$$

The result then follows from the Landau–Siegel theorem and the bound $L(1, \chi_D) \ll \log |D|$. □

6.2.2 Genus characters

The character group $\widehat{\mathrm{Gen}_K}$ of Gen_K is the group of real characters of Cl_K^+. These genus characters are indexed by unordered pairs of coprime fundamental discriminants $d_1, d_2 \in \mathbb{Z}$ satisfying $d_1 d_2 = D$. To each pair d_1, d_2, we let $\chi = \chi_{d_1,d_2}$ denote the genus character corresponding to d_1, d_2: this is a real character of the narrow class group Cl_K^+ that extends multiplicatively to all nonzero fractional ideals via

$$\chi(\mathfrak{p}) := \begin{cases} \chi_{d_1}(N(\mathfrak{p})) & \text{if } (N(\mathfrak{p}), d_1) = 1, \\ \chi_{d_2}(N(\mathfrak{p})) & \text{if } (N(\mathfrak{p}), d_2) = 1, \end{cases}$$

for any prime ideals $\mathfrak{p} \nmid \mathfrak{d}_K$, where χ_{d_1}, χ_{d_2} are the primitive real Dirichlet characters modulo d_1, d_2 respectively. In particular, χ is a quadratic character unless either d_1 or d_2 is 1, in which case it is the trivial character.

Lemma 5.6 *For any narrow ideal classes $A_1, A_2 \in \mathrm{Cl}_K^+$, we have that*

$$\frac{1}{2^{\omega(|D|)-1}} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \chi(A_1 A_2) = \begin{cases} 1 & \text{if } A_2 \in A_1 (\mathrm{Cl}_K^+)^2, \\ 0 & \text{otherwise.} \end{cases}$$

Proof This is character orthogonality for finite abelian groups. □

We abuse notation and write G_K for an element in the coset of Cl_K^+ corresponding to the genus G_K. This allows us to write

$$W_{G_K(z_A),f} = \frac{1}{2^{\omega(-D)-1}} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \chi(G_K) \sum_{A \in \mathrm{Cl}_K} \chi(A) f(z_A),$$

$$W_{G_K(\mathcal{C}_A),f} = \frac{1}{2^{\omega(D)-1}} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \chi(G_K) \sum_{A \in \mathrm{Cl}_K^+} \chi(A) \int_{\mathcal{C}_A} f(z)\, ds,$$

$$W_{G_K(\Gamma_A \backslash \mathcal{N}_A),f} = \frac{1}{2^{\omega(D)-1}} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \chi(G_K) \sum_{A \in \mathrm{Cl}_K^+} \chi(A) \int_{\Gamma_A \backslash \mathcal{N}_A} f(z)\, d\mu(z),$$

and analogous identities for $W_{G_K(z_A),\infty}(t)$, $W_{G_K(\mathcal{C}_A),\infty}(t)$, and $W_{G_K(\Gamma_A\backslash\mathcal{N}_A),\infty}(t)$. This has the advantage that we are able to show in each case that the square of the sum over $A \in \mathrm{Cl}_K^+$ is essentially equal to a product of L-functions.

6.2.3 Maaß form Weyl sums

Lemma 5.7 *We have that*

$$\left|W_{G_K(z_A),f}\right|^2 \ll \sqrt{-D} \sum_{\chi\in\widehat{\mathrm{Gen}_K}} \frac{L\left(\frac{1}{2}, f\otimes\chi_{d_1}\right) L\left(\frac{1}{2}, f\otimes\chi_{d_2}\right)}{L\left(1, \mathrm{sym}^2 f\right)},$$

$$\left|W_{G_K(\mathcal{C}_A),f}\right|^2 \ll \sqrt{\frac{D}{\frac{1}{4}+t_f^2}} \sum_{\substack{\chi\in\widehat{\mathrm{Gen}_K}\\ d_1,d_2>0}} \frac{L\left(\frac{1}{2}, f\otimes\chi_{d_1}\right) L\left(\frac{1}{2}, f\otimes\chi_{d_2}\right)}{L\left(1, \mathrm{sym}^2 f\right)},$$

$$\left|W_{G_K(\Gamma_A\backslash\mathcal{N}_A),f}\right|^2 \ll \sqrt{\frac{D}{\left(\frac{1}{4}+t_f^2\right)^3}} \sum_{\substack{\chi\in\widehat{\mathrm{Gen}_K}\\ d_1,d_2<0}} \frac{L\left(\frac{1}{2}, f\otimes\chi_{d_1}\right) L\left(\frac{1}{2}, f\otimes\chi_{d_2}\right)}{L\left(1, \mathrm{sym}^2 f\right)}.$$

Proof For $\chi = \chi_{d_1,d_2}$ and even $f \in \mathcal{B}_0(\Gamma)$, it is shown in [11, Theorem 4 and Equation (5.17)] that the quantity

$$\begin{cases} \left|\sum_{A\in\mathrm{Cl}_K} \chi(A)\frac{4\sqrt{\pi}}{w_K} f(z_A)\right|^2 & \text{if } d_1 d_2 < 0, \\ \left|\sum_{A\in\mathrm{Cl}_K^+} \chi(A)\int_{\mathcal{C}_A} f(z)\, ds\right|^2 & \text{if } d_1, d_2 > 0, \\ \left|\sum_{A\in\mathrm{Cl}_K^+} \chi(A)\frac{\frac{1}{4}+t_f^2}{2}\int_{\Gamma_A\backslash\mathcal{N}_A} f(z)\, d\mu(z)\right|^2 & \text{if } d_1, d_2 < 0 \end{cases} \tag{6.8}$$

is equal to

$$\frac{1}{2}\frac{\Lambda\left(\frac{1}{2}, f\otimes\chi_{d_1}\right)\Lambda\left(\frac{1}{2}, f\otimes\chi_{d_2}\right)}{\Lambda\left(1, \mathrm{sym}^2 f\right)}.$$

Here we recall the definition (2.7) of the completed L-function, and in particular that this includes the conductor. This identity also holds when f is odd, because in this case $L(1/2, f\otimes\chi_d) = 0$. Finally, it is also shown that

$$\sum_{A\in\mathrm{Cl}_K^+} \chi(A)\int_{\Gamma_A\backslash\mathcal{N}_A} f(z)\, d\mu(z)$$

vanishes if $d_1, d_2 > 0$, and similarly

$$\sum_{A \in \mathrm{Cl}_K^+} \chi(A) \int_{\mathcal{C}_A} f(z)\, ds$$

vanishes if $d_1, d_2 < 0$. The result then follows from the Cauchy-Schwarz inequality and Stirling's approximation. □

Remark 5.9 The terms (6.8) can be viewed as toric integrals in the sense of [40, Sect. 2.2.1], and these can be generalised to involve Hecke Größencharaktere χ of K that are not necessarily genus characters. The resulting toric integral in this generalised setting will essentially be equal to the completed Rankin–Selberg L-function $\Lambda(1/2, f \otimes g_\chi)$, where g_χ denotes the automorphic induction of the Hecke Größencharakter χ to a Maaß newform g_χ. When χ is a genus character χ_{d_1,d_2}, this Rankin–Selberg L-function factorises as $\Lambda(1/2, f \otimes \chi_{d_1})\Lambda(1/2, f \otimes \chi_{d_2})$, while the case of χ being an ideal class character of an imaginary quadratic field K and its applications towards equidistribution of Heegner points in conjugates of $\Gamma\backslash\mathbb{H}$ in $\Gamma_0(q)\backslash\mathbb{H}$ is investigated in [37].

6.2.4 Eisenstein series Weyl sums

Lemma 5.10 *We have that*

$$\left|W_{G_K(z_A),\infty}(t)\right|^2 \ll \sqrt{-D} \sum_{\chi \in \widehat{\mathrm{Gen}}_K} \left| \frac{L\left(\frac{1}{2}+it, \chi_{d_1}\right) L\left(\frac{1}{2}+it, \chi_{d_2}\right)}{\zeta(1+2it)} \right|^2,$$

$$\left|W_{G_K(\mathcal{C}_A),\infty}(t)\right|^2 \ll \sqrt{\frac{D}{\frac{1}{4}+t^2}} \sum_{\substack{\chi \in \widehat{\mathrm{Gen}}_K \\ d_1,d_2>0}} \left| \frac{L\left(\frac{1}{2}+it, \chi_{d_1}\right) L\left(\frac{1}{2}+it, \chi_{d_2}\right)}{\zeta(1+2it)} \right|^2,$$

$$\left|W_{G_K(\Gamma_A\backslash\mathcal{N}_A),\infty}(t)\right|^2 \ll \sqrt{\frac{D}{\left(\frac{1}{4}+t^2\right)^3}} \sum_{\substack{\chi \in \widehat{\mathrm{Gen}}_K \\ d_1,d_2<0}} \left| \frac{L\left(\frac{1}{2}+it, \chi_{d_1}\right) L\left(\frac{1}{2}+it, \chi_{d_2}\right)}{\zeta(1+2it)} \right|^2.$$

Proof This follows from [11, Theorem 3], akin to the proof of Lemma 5.7. □

6.3 Bounds for the variances

Proof of Proposition 5.2 For $R \asymp (-D)^{-\delta}$, $\mathrm{Var}\,(G_K(z_A); R)$ is bounded by a constant multiple dependent on ε of

$$(-D)^{-\frac{1}{2}+\varepsilon} \sum_{\chi \in \widehat{\mathrm{Gen}}_K} \sum_{0<t_f<2(-D)^\delta} \frac{L\left(\frac{1}{2}, f \otimes \chi_{d_1}\right) L\left(\frac{1}{2}, f \otimes \chi_{d_2}\right)}{L\left(1, \mathrm{sym}^2 f\right)}$$

$$+ (-D)^{-\frac{1}{2}+3\delta+\varepsilon} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \sum_{t_f \geq 2(-D)^{\delta}} \frac{L\left(\frac{1}{2}, f \otimes \chi_{d_1}\right) L\left(\frac{1}{2}, f \otimes \chi_{d_2}\right)}{t_f^3 L\left(1, \mathrm{sym}^2 f\right)}$$

$$+ (-D)^{-\frac{1}{2}+\varepsilon} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \int_0^{2(-D)^{\delta}} \frac{\left|L\left(\frac{1}{2}+it, \chi_{d_1}\right)\right|^2 \left|L\left(\frac{1}{2}+it, \chi_{d_2}\right)\right|^2}{|\zeta(1+2it)|^2} \, dt$$

$$+ (-D)^{-\frac{1}{2}+3\delta+\varepsilon} \sum_{\chi \in \widehat{\mathrm{Gen}_K}} \int_{2(-D)^{\delta}}^{\infty} \frac{\left|L\left(\frac{1}{2}+it, \chi_{d_1}\right)\right|^2 \left|L\left(\frac{1}{2}+it, \chi_{d_2}\right)\right|^2}{t^3 |\zeta(1+2it)|^2} \, dt$$

via Proposition 5.4 and Lemmata 5.5, 5.7, and 5.10; an analogous bound also holds for $\mathrm{Var}(G_K(\mathcal{C}_A); R)$. Making use of the generalised Lindelöf hypothesis in each expression and using the Weyl law yields Proposition 5.2. □

For unconditional results, we make use of the following bounds.

Lemma 5.11 ([27, Theorem]) *For $T \gg 1$,*

$$\sum_{T \leq t_f \leq T+1} \frac{L\left(\frac{1}{2}, f\right)^3}{L(1, \mathrm{sym}^2 f)} \ll_{\varepsilon} T^{1+\varepsilon},$$

$$\int_T^{T+1} \frac{\left|\zeta\left(\frac{1}{2}+it\right)\right|^6}{|\zeta(1+2it)|^2} \, dt \ll_{\varepsilon} T^{1+\varepsilon}.$$

Lemma 5.12 ([48, Theorem 1.1]) *For odd fundamental discriminants $D \neq 1$ and $T \gg 1$,*

$$\sum_{T \leq t_f \leq T+1} \frac{L\left(\frac{1}{2}, f \otimes \chi_D\right)^3}{L(1, \mathrm{sym}^2 f)} \ll_{\varepsilon} (|D|T)^{1+\varepsilon},$$

$$\int_T^{T+1} \frac{\left|L\left(\frac{1}{2}+it, \chi_D\right)\right|^6 \, dt}{|\zeta(1+2it)|^2} \ll_{\varepsilon} (|D|T)^{1+\varepsilon}.$$

Proof of Proposition 5.3 We bound the variance by breaking up into ranges as in the proof of Proposition 5.2. Instead of applying the generalised Lindelöf hypothesis, we use the generalised Hölder inequality with exponents (3, 3, 3). Via the bounds in Lemmata 5.11 and 5.12, together with the Weyl law, we obtain the result. □

6.4 Representations of integers by indefinite ternary quadratic forms

We briefly describe how the results in this section can be interpreted in terms of indefinite ternary quadratic forms. For simplicity, we only discuss the case of negative discriminant and summing over all genera; for positive discriminant, a detailed presentation can be found in [12, Sect. 2].

Consider the indefinite ternary quadratic form

$$Q(a, b, c) = b^2 - 4ac.$$

We are interested in the level sets

$$V_{Q,D}(\mathbb{Z}) := \left\{(a, b, c) \in \mathbb{Z}^3 : b^2 - 4ac = D\right\},$$

where $D < 0$ is a fundamental discriminant; these sets parametrise the different ways that the integer D can be represented by the ternary quadratic form Q. The normalised level set $\mathscr{G}_D := (-D)^{-1/2} V_{Q,D}(\mathbb{Z})$ lies inside the two-sheeted hyperboloid

$$V_{Q,-1}(\mathbb{R}) := \left\{(a, b, c) \in \mathbb{R}^3 : b^2 - 4ac = -1\right\}.$$

It is natural to ask whether the normalised level sets $\mathscr{G}_D$ cover $V_{Q,-1}(\mathbb{R})$ randomly as D tends to $-\infty$ along fundamental discriminants. Each level set $V_{Q,D}(\mathbb{Z})$ is countably infinite, and $V_{Q,-1}(\mathbb{R})$ is isomorphic to $\mathbb{C}\backslash\mathbb{R}$, which is not of finite volume, so one cannot immediately rephrase this random covering as equidistribution.

On the other hand, the group

$$\mathrm{SO}_Q(\mathbb{Z}) := \left\{A \in \mathrm{SL}_3(\mathbb{Z}) : Q(Ax) = Q(x) \text{ for all } x = (a, b, c) \in \mathbb{Z}^3\right\}$$

acts transitively on $V_{Q,D}(\mathbb{Z})$, and the quotient space $\mathrm{SO}_Q(\mathbb{Z})\backslash\mathscr{G}_D$ is finite for all fundamental discriminants D, with cardinality equal to h_K. Moreover, $\mathrm{SO}_Q(\mathbb{Z})$ is a discrete subgroup of $\mathrm{SO}_Q(\mathbb{R})$ of finite covolume, and $V_{Q,-1}(\mathbb{R}) \cong \mathrm{SO}_Q(\mathbb{R})/K$ with K equal to the maximal compact subgroup of $\mathrm{SO}_Q(\mathbb{R})$, and so the space $\mathrm{SO}_Q(\mathbb{Z})\backslash V_{Q,-1}(\mathbb{R})$ is of finite volume.

Thus to ask whether the normalised level sets $\mathscr{G}_D$ randomly cover $V_{Q,-1}(\mathbb{R})$ can be rephrased as asking whether the finite sets $\mathrm{SO}_Q(\mathbb{Z})\backslash\mathscr{G}_D$ equidistribute in the finite volume space $\mathrm{SO}_Q(\mathbb{Z})\backslash V_{Q,-1}(\mathbb{R})$. This has a positive answer by naturally realising this result in terms of the equidistribution of Heegner points on $\Gamma\backslash\mathbb{H}$, as proved by Duke [10, Theorem 1]. Indeed, the fact that Q is indefinite implies that SO_Q is isomorphic to the split special orthogonal group $\mathrm{SO}_{1,2}$, and we have the accidental isomorphism $\mathrm{SO}_{1,2} \cong \mathrm{PGL}_2$, while $K \cong \mathrm{SO}_2(\mathbb{R})$. From this, we see that $\mathrm{SO}_Q(\mathbb{Z})\backslash V_{Q,-1}(\mathbb{R}) \cong \mathrm{PGL}_2(\mathbb{Z})\backslash\mathrm{PGL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R}) \cong \Gamma\backslash\mathbb{H}$, while $\mathrm{SO}_Q(\mathbb{Z})\backslash\mathscr{G}_D$ is naturally identified with the set of Heegner points $\{z_A \in \Gamma\backslash\mathbb{H} : A \in \mathrm{Cl}_K\}$.

With this reinterpretation in mind, we now see that Proposition 5.2 implies that under the assumption of the generalised Lindelöf hypothesis, almost every shrinking ball of radius $R \asymp (-D)^{-\delta}$ with $0 < \delta < 1/4$ in $\mathrm{SO}_Q(\mathbb{Z})\backslash V_{Q,-1}(\mathbb{R})$ contains a normalised equivalence class of points $(a, b, c) \in \mathbb{Z}^3$ that represent the integer D by the indefinite ternary quadratic form $Q(a, b, c) = b^2 - 4ac$. This complements [4, Theorem 1.8], where the analogous result is proved for the definite ternary quadratic form $Q(a, b, c) = a^2 + b^2 + c^2$.

Acknowledgements The author thanks Peter Sarnak for suggesting this problem and many helpful discussions on this topic, as well as Matt Young for useful feedback.

References

1. Berry, M.V.: Regular and irregular semiclassical wavefunctions. J. Phys. A: Math. Gen. **10**(12), 2083–2091 (1977). https://doi.org/10.1088/0305-4470/10/12/016
2. Blomer, V., Milićević, D.: The second moment of twisted modular L-functions. Geom. Funct. Anal. **25**(2), 453–516 (2015). https://doi.org/10.1007/s00039-015-0318-7
3. Bourgain, J.: Decoupling, exponential sums and the Riemann Zeta function. J. Am. Math. Soc. **30**(1), 205–224 (2017). https://doi.org/10.1090/jams/860
4. Bourgain, J., Rudnick, Z., Sarnak, P.: Spatial Statistics for Lattice Points on the Sphere I: Individual Results. Bull. Iran. Math. Soc. **43**(4), 361–38 (2017). http://bims.iranjournals.ir/article_1169.html
5. Buttcane, J., Khan, R.: A mean value of triple product L-functions. Math. Z. **285**(1), 565–591 (2017). https://doi.org/10.1007/s00209-016-1721-y
6. Buttcane, J., Khan, R.: On the fourth moment of Hecke Maass forms and the random wave conjecture. Compos. Math. **153**(7), 1479–1511 (2017). https://doi.org/10.1112/S0010437X17007199
7. Chamizo, F.: Some applications of large sieve in Riemann surfaces. Acta Arith. **77**(4), 315–337 (1996). https://doi.org/10.4064/aa-77-4-315-337
8. Chelluri, T.: Equidistribution of the Roots of Quadratic Congruences, Ph.D. Thesis, Rutgers The State University of New Jersey, New Brunswick, (2004)
9. Djanković, G., Khan, R.: A conjecture for the regularized fourth moment of Eisenstein series. J. Num. Theory **182**, 236–257 (2018). https://doi.org/10.1016/j.jnt.2017.06.012
10. Duke, W.: Hyperbolic distribution problems and half-integral weight maass forms. Invent. Math. **92**(1), 73–90 (1988). https://doi.org/10.1007/BF01393993
11. Duke, W., Imamoglu, Ö., Tóth, Á.: Geometric Invariants for Real Quadratic Fields. Ann. Math. 184(3), 949–990 (2016). https://doi.org/10.4007/annals.2016.184.3.8
12. Einsiedler, M., Lindenstrauss, E., Michel, P., Venkatesh, A.: The distribution of closed geodesics on the modular surface, and Duke's theorem. L'Enseignement Mathématique **58**, 249–313 (2012). https://doi.org/10.4171/LEM/58-3-2
13. Ellenberg, J.S., Michel, P., Venkatesh, A.: Linnik's Ergodic Method and the Distribution of Integer Points on Spheres. In: Automorphic Representations and L-Functions. Proceedings of the International Colloquium, Mumbai 2012, editors D. Prasad, C. S. Rajan, A. Sankaranarayanan, and J. Sengupta, Hindustan Book Agency, New Delhi, 119–185 (2013)
14. Gradshteyn, I.S., Ryzhik, I.M.: In: Jeffrey, A., Zwillinger, D. (eds.) Table of Integrals, Series, and Products, 7th edn. Academic Press, Burlington (2007)
15. Granville, A., Wigman, I.: Planck-scale mass equidistribution of toral laplace eigenfunctions. Commun. Math. Phys. **355**(2), 767–802 (2017). https://doi.org/10.1007/s00220-017-2953-3
16. Han, X.: Small scale quantum ergodicity in negatively curved manifolds. Nonlinearity **28**(9), 3263–3288 (2015). https://doi.org/10.1088/0951-7715/28/9/3263
17. Han, X.: Small scale quantum ergodicity of random eigenbases. Commun. Math. Phys. **349**(1), 425–440 (2017). https://doi.org/10.1007/s00220-016-2597-8
18. Han, X., Tacy, M.: Equidistribution of Random Waves on Small Balls. Preprint (2016), 13 pages. arXiv:1611.05983
19. Hejhal, D.A.: On Eigenfunctions of the Laplacian for Hecke Triangle Groups. In: Hejhal, Dennis A., Friedman, Joel, Gutzwiller, Martin C., Odlyzko, Andrew M. (eds.) Emerging Applications of Number Theory, pp. 291–315. The IMA Volumes in Mathematics and Its Applications **109**, Springer–Verlag, New York (1999). https://doi.org/10.1007/978-1-4612-1544-8_11
20. Hejhal, D.A., Rackner, B.N.: On the topography of Maass waveforms for $\mathrm{PSL}(2, \mathbb{Z})$. Exp. Math. **1**(4), 275–305 (1992). https://doi.org/10.1080/10586458.1992.10504562

21. Hejhal, D.A., Strömbergsson, A.: On quantum chaos and Maass waveforms of CM-type. Found. Phys. **31**(3), 519–533 (2001). https://doi.org/10.1023/A:1017521729782
22. Hezari, H., Rivière, G.: L^p norms, nodal sets, and quantum ergodicity. Adv. Math. **290**, 938–966 (2016). https://doi.org/10.1016/j.aim.2015.10.027
23. Hezari, H., Rivière, G.: Quantitative equidistribution properties of toral eigenfunctions. J. Spectr. Theory **7**(2), 471–485 (2017). https://doi.org/10.4171/JST/169
24. Hoffstein, J., Lockhart, P.: Coefficients of maass forms and the siegel zero. Ann. Math. **140**(1), 161–176 (1994). https://doi.org/10.2307/2118543
25. Hu, Y.: Triple product formula and mass equidistribution on modular curves of level N to appear. Int. Math. Res. Notices (2017), p 45 . https://doi.org/10.1093/imrn/rnw322
26. Ichino, A.: Trilinear forms and the central values of triple product L-functions. Duke Math. J. **145**(2), 281–307 (2008). https://doi.org/10.1215/00127094-2008-052
27. Ivić, A.: On sums of Hecke series in short intervals. Journal de Théorie des Nombres de Bordeaux **13**(2), 453–468 (2001). https://doi.org/10.5802/jtnb.333
28. Iwaniec, H.: Spectral methods of automorphic forms, Second Edition, Graduate Studies in Mathematics **53**, American Mathematical Society, Providence, 2002. https://doi.org/10.1090/gsm/053
29. Iwaniec, H., Kowalski, E.: Analytic number theory, American mathematical society Colloquium Publications 53, American Mathematical Society. Providence (2004). https://doi.org/10.1090/coll/053
30. Jakobson, D.: Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$. Ann. Inst. Fourier **44**(5), 1477–1504 (1994). https://doi.org/10.5802/aif.1442
31. Jutila, M.: The Fourth Moment of Central Values of Hecke Series. In: Jutila, M., Metsänkylä, T. (eds.) Number Theory: Proceedings of the Turku Symposium on Number Theory in Memory of Kustaa Inkeri, pp. 167–177. Walter de Gruyter, Berlin (2001)
32. Jutila, M.: The spectral mean square of Hecke L-functions on the critical line. Publications de l'Institut Mathématique, Nouvelle série **76**(90), 41–55 (2004). https://doi.org/10.2298/PIM0476041J
33. Jutila, M., Motohashi, Y.: Uniform bound for Hecke L-functions. Acta Math. **195**(1), 61–115 (2005). https://doi.org/10.1007/BF02588051
34. Lester, S., Matomäki, K., Radziwiłł, M.: Small Scale Distribution of Zeros and Mass of Modular Forms. Journal of the European Mathematical Society (2018), 31 pages. arXiv:1501.01292 [math.NT]
35. Lester, S., Rudnick, Z.: Small scale equidistribution of eigenfunctions on the torus. Commun. Math. Phys. **350**(1), 279–300 (2017). https://doi.org/10.1007/s00220-016-2734-4
36. Lindenstrauss, E.: Invariant measures and arithmetic quantum unique ergodicity. Ann. Math. **163**(1), 165–219 (2006). https://doi.org/10.4007/annals.2006.163.165
37. Liu, S.C., Masri, R., Young, M.P.: Subconvexity and equidistribution of heegner points in the level aspect. Compos. Math. **149**(7), 1150–1174 (2013). https://doi.org/10.1112/S0010437X13007033
38. Luo, W.: L^4-norms of the dihedral maass forms. Int. Math. Res. Notices **2014**(8), 2294–2304 (2014). https://doi.org/10.1093/imrn/rns298
39. Luo, W., Sarnak, P.: Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2$. Publications Mathématiques de l'Institut des Hautes Études Scientifiques **81**(1), 207–237 (1995). https://doi.org/10.1007/BF02699377
40. Michel, P., Venkatesh, A.: Equidistribution, L-Functions and Ergodic Theory: On Some Problems of Yu. Linnik. In: Proceedings of the International Congress of Mathematicians, Madrid 2006 **II**, editors Marta Sanz-Solé, Javier Soria, Juan Luis Varona, and Joan Verdera, European Mathematical Society, Zürich, (2006), 421–457. http://www.icm2006.org/proceedings/Vol_II/contents/ICM_Vol_2_19.pdf
41. Milićević, D.: Large values of eigenfunctions on arithmetic hyperbolic surfaces. Duke Math. J. **155**(2), 365–401 (2010). https://doi.org/10.1215/00127094-2010-058
42. Nelson, Paul D., Pitale, Ameya, Saha, Abhishek: Bounds for rankin-selberg integrals and quantum unique ergodicity for powerful levels. J. Am. Math. Soc. **27**(1), 147–191 (2014). https://doi.org/10.1090/S0894-0347-2013-00779-1
43. Sarnak, P.: Spectra of hyperbolic surfaces. Bull. Am. Math. Soc. **40**(4), 441–478 (2003). https://doi.org/10.1090/S0273-0979-03-00991-1
44. Soundararajan, K.: Quantum unique ergodicity for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. Ann. Math. **172**(2), 1529–1538 (2010). https://doi.org/10.4007/annals.2010.172.1529
45. Spinu, F.: The L^4 Norm of the Eisenstein Series, Ph.D. Thesis, Princeton University, (2003). http://www.math.jhu.edu/~fspinu/math/thesis.pdf
46. Watson, T.C.: Rankin Triple Products and Quantum Chaos, Ph.D. Thesis, Princeton University, (2002) (revised 2008). arXiv:0810.0425 [math.NT]

47. Young, M.P.: The quantum unique ergodicity conjecture for thin sets. Adv. Math. **286**, 958–1016 (2016). https://doi.org/10.1016/j.aim.2015.09.013
48. Young, M.P.: Weyl-type hybrid subconvexity bounds for twisted L-functions and Heegner points on shrinking sets. J. Eur. Math. Soc. **19**(5), 1545–1576 (2017). https://doi.org/10.4171/JEMS/699
49. Young, M.P.: Explicit Calclulations with Eisenstein Series. preprint (2017), 37 pages. arXiv:1710.03624 [math.NT]
50. Zagier, D.: The Rankin–Selberg Method for Automorphic Functions which Are Not of Rapid Decay. Journal of the Faculty of Science, the University of Tokyo. Sect. 1 A, Mathematics **28**(3), 415–437 (1982). http://hdl.handle.net/2261/6300