EQUIDISTRIBUTION OF CUSP FORMS IN THE LEVEL ASPECT

PAUL D. NELSON

ABSTRACT. Let f traverse a sequence of classical holomorphic newforms of fixed weight and increasing squarefree level $q \to \infty$. We prove that the pushforward of the mass of f to the modular curve of level 1 equidistributes with respect to the Poincaré measure.

Our result answers affirmatively the squarefree level case of a conjecture spelled out by Kowalski, Michel and Vanderkam (2002) in the spirit of a conjecture of Rudnick and Sarnak (1994).

Our proof follows the strategy of Holowinsky and Soundararajan (2008) who show that newforms of level 1 and large weight have equidistributed mass. The new ingredients required to treat forms of fixed weight and large level are an adaptation of Holowinsky's reduction of the problem to one of bounding shifted sums of Fourier coefficients (which on the surface makes sense only in the large weight limit), an evaluation of the p-adic integral needed to extend Watson's formula to the case of three newforms where the level of one divides but need not equal the common squarefree level of the other two, and some additional technical work in the problematic case that the level has many small prime factors.

CONTENTS

1. Introduction 2
1.1. Statement of result 2
1.2. Comparison with Holowinsky-Soundararajan (2008) 3
1.3. Plan for the paper 5
1.4. Notation and conventions 5
1.5. Weyl's criterion 6
1.6. Acknowledgements 6
2. Background on automorphic forms 6
2.1. Holomorphic newforms 6
2.2. Maass eigencuspforms 7
2.3. Eisenstein series 8
2.4. Incomplete Eisenstein series 8
3. Main estimates 8
3.1. Reduction to shifted sums 9
3.2. Bounds for individual shifted sums 12
3.3. Bounds for sums of shifted sums 14
4. An extension of Watson's formula 16
5. Proof of Theorem 1.3 20
References 21

1. Introduction

1.1. **Statement of result.** A basic problem in modern number theory and the analytic theory of modular forms is to understand the limiting behavior of modular forms in families. Let $f : \mathbb{H} \to \mathbb{C}$ be a classical holomorphic newform of weight k and level q. The *mass* of f is the finite measure $d\nu_f = |f(z)|^2 y^{k-2}\, dx\, dy$ ($z = x+iy$) on the modular curve $Y_0(q) = \Gamma_0(q)\backslash\mathbb{H}$. In a recent breakthrough, Holowinsky and Soundararajan [14] proved that newforms of large weight k and fixed level $q = 1$ have equidistributed mass, answering affirmatively a natural variant[1] of the *quantum unique ergodicity* conjecture of Rudnick and Sarnak [30].

Theorem 1.1 (Mass equidistribution for $\mathrm{SL}(2,\mathbb{Z})$ in the weight aspect)**.** *Let f traverse a sequence of newforms of increasing weight $k \to \infty$ and fixed level $q = 1$. Then the mass ν_f equidistributes*[2] *with respect to the Poincaré measure $d\mu = y^{-2}\, dx\, dy$ on the modular curve $Y_0(q)$.*

Kowalski, Michel and Vanderkam [20, Conj 1.5] formulated an analogue of the Rudnick-Sarnak conjecture in which the roles of the parameters k and q are reversed: they conjectured that the masses of newforms of fixed weight and large level q are equidistributed amongst the fibers of the canonical projection $\pi_q : Y_0(q) \to Y_0(1)$ in the following sense.

Conjecture 1.2 (Mass equidistribution for $\mathrm{SL}(2,\mathbb{Z})$ in the level aspect)**.** *Let f traverse a sequence of newforms of fixed weight and increasing level $q \to \infty$. Then the pushforward $\mu_f := \pi_{q*}(\nu_f)$ of the mass of f to $Y_0(1)$ equidistributes with respect to μ.*

Kowalski, Michel and Vanderkam remark that Conjecture 1.2 follows in the special case of dihedral forms from their subconvex bounds for Rankin-Selberg L-functions modulo an unestablished extension of Watson's formula [42], which is now known by Theorem 4.1 of this paper. Recently Koyama [21], following the method of Luo and Sarnak [23], proved the analogue of Conjecture 1.2 for unitary Eisenstein series of increasing prime level by reducing the problem to known subconvex bounds for automorphic L-functions of degree two.

Our aim in this paper is to establish the squarefree level case of Conjecture 1.2. Our result is the first of its kind for non-dihedral cusp forms.

Theorem 1.3 (Mass equidistribution for $\mathrm{SL}(2,\mathbb{Z})$ in the squarefree level aspect)**.** *Let f traverse a sequence of newforms of fixed weight and increasing squarefree level $q \to \infty$. Then μ_f equidistributes with respect to μ.*

Remark 1.4. Our extension (Theorem 4.1) of Watson's formula [42] shows that Theorem 1.3 would follow immediately from an appropriate generalization of the Riemann hypothesis, so one can view Theorem 1.3 as an unconditionally known consequence of a central unresolved conjecture.

Remark 1.5. One cannot relax entirely the restriction of Theorem 1.3 to newforms, since for instance a cusp form of level 1 may be regarded as an oldform of arbitrary level $q > 1$.

Remark 1.6. Rudnick [29] showed that Theorem 1.1 implies that the zeros of newforms of level 1 and weight $k \to \infty$ equidistribute on $Y_0(1)$. Soundararajan [40] asks whether there is an analogue of Rudnick's result for newforms of large level. We do not know whether such an analogue exists and highlight here one of the difficulties in adapting Rudnick's method. Let f be a newform of weight k and level q, let $\mathcal{Z}$ be the left $\Gamma_0(q)$-multiset of zeros of f in $\mathbb{H}$ and let $\mathcal{Z}_1$ be the left Γ-multiset ($\Gamma = \mathrm{PSL}(2,\mathbb{Z})$) obtained by summing the images of $\mathcal{Z}$ under coset representatives for $\Gamma(1)/\Gamma_0(q)$. We ask: does $\Gamma\backslash\mathcal{Z}_1$ equidistribute on $Y_0(1)$ as

[1] as spelled out by Luo and Sarnak [24]; we refer to Sarnak [31, 32] and the references in [14] for further discussion.

[2] We say that a sequence of finite Radon measures μ_j on a locally compact Hausdorff space X *equidistributes* with respect to some fixed finite Radon measure μ if for each function $\phi \in C_c(X)$ we have $\mu_j(\phi)/\mu_j(1) \to \mu(\phi)/\mu(1)$ as $j \to \infty$, here and always identifying a measure μ with the corresponding linear functional $\phi \mapsto \mu(\phi) := \int_X \phi\, d\mu$ on the space $C_c(X)$.

$q \to \infty$? Following Rudnick, one may show for $\phi \in C_c^\infty(\mathbb{H})$ and $\Phi(z) = \sum_{\gamma \in \Gamma} \phi(\gamma z)$ that

$$\frac{12}{k\psi(q)} \sum_{z \in \Gamma \backslash \mathcal{Z}_1} \frac{\Phi(z)}{\# \operatorname{Stab}_\Gamma(z)} = \int_{\Gamma \backslash \mathbb{H}} \Phi \, dV + \int_{\Gamma \backslash \mathbb{H}} \frac{\pi_{q*}(\log \nu_f)}{k\psi(q)} \Delta \Phi \, dV, \tag{1}$$

where $\psi(q) = [\Gamma(1) : \Gamma_0(q)]$, $\Delta = y^2(\partial_x^2 + \partial_y^2)$ is the hyperbolic Laplacian, and dV is the hyperbolic probability measure on $\Gamma \backslash \mathbb{H}$; the formula (1) follows by some elementary manipulations of the identity $\int_{\mathbb{H}} \log |z - z_0| \Delta \phi(z) y^{-2} \, dx \, dy = 2\pi \phi(z_0)$, which holds for any $z_0 \in \mathbb{H}$ and follows from Green's identities. Since the total number of inequivalent zeros is $\# \Gamma \backslash \mathcal{Z}_1 = \# \Gamma_0(q) \backslash \mathcal{Z} \sim k\psi(q)/12$ [36, §2], the first term on the right hand side of (1) may be regarded as a main term, the second as an error term that one would like to show tends to 0. An important step towards adapting Rudnick's method would be to rule out the possibility that $\pi_{q*}(\log \nu_f)/k\psi(q)$ tends to $-\infty$ uniformly on compact subsets as $q \to \infty$. The difficulty in doing so is that Theorem 1.3 does not seem to preclude the masses ν_f from being very small somewhere within each fiber of the projection $Y_0(q) \to Y_0(1)$. There are further difficulties in adapting Rudnick's method that we shall not mention here.

Remark 1.7. Lindenstrauss [22] and Soundararajan [39] proved that Maass eigencuspforms of fixed level q and large Laplace eigenvalue $\lambda \to \infty$ have equidistributed mass. We ask: do Maass newforms of large level $q \to \infty$ (with λ taken to lie in a fixed subinterval of $[1/4, +\infty]$, say) satisfy the natural analogue of Conjecture 1.2? An affirmative answer to this question would follow from a suitable generalization of the Riemann hypothesis (at least for q squarefree, as in Remark 1.4), but appears beyond the reach of our methods because the Ramanujan conjecture is not known for Maass forms (compare with [14, p2]).

Remark 1.8. With minor modifications our arguments establish the following stronger hybrid equidistribution result: for a newform f of possibly *varying* weight k and squarefree level q, the measures $\mu_f = \pi_{q*}(\nu_f)$ equidistribute as $qk \to \infty$. In this paper we fix the weight k only to simplify the exposition.

Remark 1.9. With minor modifications our arguments should extend to the general case of not necessarily squarefree levels q as soon as an appropriate extension of Watson's formula is worked out. However, we shall invoke the assumption that the level q is squarefree whenever doing so simplifies the exposition. The parts of our argument that require modification to treat the general case are Lemmas 3.4, 3.15 and 4.4. One should be able to generalize Lemmas 3.4 and 3.15 using that for any level q the cusps of $\Gamma_0(q)$ fall into classes indexed by the divisors d of q consisting of $\phi(\gcd(d, q/d))$ cusps of width $d/\gcd(d, q/d)$. To generalize 4.4, one must compute a p-adic integral involving matrix coefficients of supercuspidal representations of $\mathrm{GL}(2, \mathbb{Q}_p)$. We plan to consider this generalization in future work.

1.2. **Comparison with Holowinsky-Soundararajan (2008).** Our proof of Theorem 1.3, which concerns the equidistribution of measures associated to newforms in the *level aspect*, is modeled after the Holowinsky-Soundararajan proof of Theorem 1.1, which concerns equidistribution in the *weight aspect*. Holowinsky [15] and Soundararajan [38] show by independent arguments that all forms not belonging to some (small) exceptional set have equidistributed mass in the large weight limit; in their joint work, Holowinsky-Soundararajan [14] then prove Theorem 1.1 by showing that the intersection of the exceptional sets of forms not covered by either of their approaches is empty.

The independent arguments of Holowinsky and of Soundararajan each consist of a preliminary "reduction" step, in which they relate the equidistribution problem to one of showing that certain expressions are small, followed by an "analysis" step in which they bound such expressions. For Holowinsky, the reduction step relates the problem to bounding sums roughly of the form $\sum_{n \leq k} \lambda_f(n) \lambda_f(n + l)$ with $l \neq 0$, where $\lambda_f(n)$ is the nth Fourier coefficient of the newform f of weight k and level 1; the analysis step is then an application of bounds he develops for such sums. For Soundararajan, the reduction step is afforded by Watson's formula, which relates the "Weyl periods" $\mu_f(\phi)$ for ϕ a fixed Maass form of level 1 to the central value of the triple product L-function $L(\frac{1}{2}, \phi \times f \times f)$; the analysis step is provided by the "weak subconvex" bounds that he develops for the central values of quite general L-functions.

The primary difficulty in adapting the method of Holowinsky and Soundararajan to the level aspect is that, for quite different reasons, the reduction steps in their independent arguments do not apply without essential modification. In Soundararajan's reduction, the reason is simple: Watson's formula applies to $\mu_f(\phi)$ only when f and ϕ are newforms of the *same* squarefree level, while in the level aspect the relevant Weyl periods are those for which f has large level and ϕ has level 1. We extend Watson's formula appropriately in Theorem 4.1 by computing (Lemma 4.4) a p-adic integral arising in Ichino's general formula [16], specifically the integral

$$\int_{g \in \mathrm{PGL}_2(\mathbb{Q}_p)} \frac{\langle g \cdot \phi_p, \phi_p \rangle}{\langle \phi_p, \phi_p \rangle} \frac{\langle g \cdot f_p, f_p \rangle}{\langle f_p, f_p \rangle} \frac{\langle g \cdot f_p, f_p \rangle}{\langle f_p, f_p \rangle} \, dg \tag{2}$$

where ϕ_p and f_p are local components at p of the adelizations of ϕ and f, and $\langle , \rangle$ denotes a G-invariant hermitian pairing. The crucial case for us is when p divides the squarefree level q of the newform f, so that ϕ_p lives in a spherical representation of $\mathrm{PGL}_2(\mathbb{Q}_p)$ and f_p in a special representation.

On the other hand, Holowinsky's reduction seems to rely on the asymptotic analysis of certain integrals that have clear behavior only in the large weight limit. We nevertheless reduce the problem of equidistribution in the level aspect to one of bounding (smoothed) sums roughly of the form

$$\sum_{d|q} \sum_{n \leq dY} \lambda_f(n) \lambda_f(n + dl) \tag{3}$$

with $l \neq 0$, where q is the squarefree level of the newform f and Y is a real parameter satisfying $1 \leq Y \ll (\log q)^{O(1)}$.

The basic idea behind this reduction, as in Holowinsky's original arguments, is to approximate $\mu_f(\phi)$ for a nice function ϕ on $Y_0(1)$ in terms of the integral of the $\Gamma(1)$-invariant measure $\phi \, d\mu_f$ taken over (a smooth truncation of) the region $\mathcal{R}$ consisting of those points $z = x + iy$ with $x \in [0, 1]$ and $y \asymp Y^{-1}$ for some (small) flexible parameter $Y \geq 1$ (Lemma 3.4); the region $\mathcal{R}$ contains roughly Y copies of a fundamental domain for $\Gamma(1)\backslash\mathbb{H}$, so that

$$\mu_f(\phi) \approx Y^{-1} \int_{\mathcal{R}} \phi \, d\mu_f.$$

At this point our arguments diverge from those of Holowinsky, since for us f lives on a covering of $Y_0(1)$. The pullback of $\mathcal{R}$ under the projection $Y_0(q) \to Y_0(1)$ is the sum of "rectangular" regions $\mathcal{R}_{\mathfrak{a}}$ indexed by the cusps $\mathfrak{a}$ of $\Gamma_0(q)$, so that

$$\int_{\mathcal{R}} \phi \, d\mu_f = \sum_{\mathfrak{a}} \int_{\mathcal{R}_{\mathfrak{a}}} \phi |f|^2 y^k \, d\mu.$$

Explicitly, we have

$$\mathcal{R}_{\mathfrak{a}} = \{z \in \mathbb{H} : x_{\mathfrak{a}} \in [0, 1], y_{\mathfrak{a}} \asymp (d_{\mathfrak{a}} Y)^{-1}\},$$

where $d_{\mathfrak{a}}$ is the width of the cusp $\mathfrak{a}$ and the proper isometry $z \mapsto z_{\mathfrak{a}} = x_{\mathfrak{a}} + i y_{\mathfrak{a}}$ is chosen so that some generator of $\mathrm{Stab}_{\Gamma_0(q)}(\mathfrak{a})$ acts on z via $z_{\mathfrak{a}} \mapsto z_{\mathfrak{a}} + 1$.

By Atkin-Lehner theory and our assumption that q is squarefree, we understand how f behaves in each region $\mathcal{R}_{\mathfrak{a}}$: its Fourier coefficients in the variable $x_{\mathfrak{a}}$ agree up to sign with its Fourier coefficients in x. Knowing this, one can show with some work that the constant term in the Fourier series for ϕ contributes the expected main term for $\mu_f(\phi)$. Since ϕ is invariant under the full modular group and in particular under $z \mapsto d_{\mathfrak{a}} z_{\mathfrak{a}}$, we have $\phi(z) = \phi(d_{\mathfrak{a}} z_{\mathfrak{a}})$, so that the non-constant terms in the Fourier series for ϕ (indexed by nonzero integers l) contribute error terms involving sums roughly of the form

$$\sum_{\mathfrak{a}} \sum_{n} \lambda_f(n) \lambda_f(n + d_{\mathfrak{a}} l) \int_0^\infty (\cdots),$$

where $\int_0^\infty(\cdots)$ is an integral with clear asymptotics only in the large weight limit. Some softer arguments (Lemma 3.12, Corollary 3.14) show that even for fixed weight, this integral essentially localizes the sum over n to the range $n \leq dY$. Since $\{d_\mathfrak{a}\} = \{d : d|q\}$ for q squarefree, we indeed reduce to bounding sums essentially of the form (3).

It remains only to estimate sums like (3). Holowinsky's bounds (Theorem 3.9) are sufficient to treat such sums when the level q has few small prime factors (for instance, if q is prime), but a refinement of his bounds (Theorem 3.10) and some additional technical work (Lemma 3.15) are needed when q has many small prime factors (for instance, if q is the product of all primes up to some bound).

1.3. **Plan for the paper.** Our paper is organized as follows. In §2 we recall some standard properties of our basic objects of study: holomorphic newforms, Maass eigencuspforms, unitary Eisenstein series and incomplete Eisenstein series.

In §3 we prove the level aspect analogue of Holowinsky's main result [15, Corollary 3]. We believe that our key contribution is recognizing the applicability of the computation occupying the latter half of the proof of Lemma 3.4; while this seems in retrospect like a natural and even obvious step, it was in fact the final of several approaches that we tried and its discovery resulted in a significant simplification of a more complicated proof that we had found earlier and do not describe in this paper. We then refine and soften Holowinsky's arguments (Theorem 3.10, Lemma 3.12, Corollary 3.14) and carry out some technical work required to treat levels q having many small prime factors (Lemma 3.15); the main idea in the latter is to reduce by a convexity argument to the case that q is the product of all primes up to some bound, in which case the prime number theorem allows one to quantify the heuristic that sufficiently many divisors d of q satisfy $\log(d) \approx \log(q)$.

In §4 we extend Watson's formula to cover the additional case that we need, building on the work of many authors [7, 27, 12, 11, 42, 16]. We keep track of various normalizations in the literature and compute a certain p-adic integral (Lemma 4.4); the computation is not difficult, but requires some careful book-keeping.

In §5 we complete the proof of Theorem 1.3 using the main results of §3 and §4. Sections 3 and 4 are independent of each other, but both depend upon the definitions, notation and facts recalled in §2.

1.4. **Notation and conventions.** Recall the standard notation for the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$, the modular group $\Gamma = \mathrm{SL}(2,\mathbb{Z})$, its congruence subgroup $\Gamma_0(q)$ consisting of those elements with lower-left entry divisible by q, the modular curve $Y_0(q) = \Gamma_0(q)\backslash\mathbb{H}$, the natural projection $\pi_q : Y_0(q) \to Y_0(1)$, the Poincaré measure $d\mu = y^{-2}\,dx\,dy$, and the stabilizer $\Gamma_\infty = \{\pm\left(\begin{smallmatrix} 1 & n \\ & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\}$ in Γ of $\infty \in \mathbb{P}^1(\mathbb{R})$. We denote a typical element of $\mathbb{H}$ as $z = x + iy$ with $x, y \in \mathbb{R}$.

There is a natural inclusion $C_c(Y_0(1)) \hookrightarrow C_c(Y_0(q))$ obtained by pulling back under the projection π_q; here C_c denotes the space of compactly-supported continuous functions. For a newform f of weight k on $\Gamma_0(q)$ the pushforward measure $d\mu_f := \pi_{q*}(|f|^2 y^k\,d\mu)$ on the modular curve $Y_0(1)$ corresponds, by definition, to the linear functional

$$\mu_f(\phi) = \int_{\Gamma_0(q)\backslash\mathbb{H}} \phi(z)|f|^2(z)y^k\,\frac{dx\,dy}{y^2} \quad \text{for } \phi \in C_c(Y_0(1)) \hookrightarrow C_c(Y_0(q)).$$

We let μ denote the standard measure on $Y_0(1)$, so that

$$\mu(\phi) = \int_{\Gamma\backslash\mathbb{H}} \phi(z)\,\frac{dx\,dy}{y^2} \quad \text{for } \phi \in C_c(Y_0(1)).$$

Since μ and μ_f are finite, they extend to the space of bounded continuous functions on $Y_0(1)$, where we shall denote also by μ and μ_f their extensions. In particular, $\mu(1)$ denotes the volume of $Y_0(1)$ and $\mu_f(1)$ the Petersson norm of f.

As is customary, we let $\varepsilon > 0$ denote a sufficiently small positive number whose precise value may change from line to line. We use the asymptotic notation $f(x,y,z) \ll_{x,y} g(x,y,z)$ to indicate that there exists a positive real $C(x,y)$, possibly depending upon x and y but not upon z, such that $|f(x,y,z)| \leq C(x,y)|g(x,y,z)|$

for all x, y and z under consideration. We write $f(x,y,z) = O_{x,y}(g(x,y,z))$ synonymously for $f(x,y,z) \ll_{x,y} g(x,y,z)$ and write $f(x,y,z) \asymp_{x,y} g(x,y,z)$ synonymously for $f(x,y,z) \ll_{x,y} g(x,y,z) \ll_{x,y} f(x,y,z)$.

1.5. **Weyl's criterion.** We conclude this introduction with a standard lemma that provides essential motivation for what follows.

The family of probability measures $\phi \mapsto \mu_f(\phi)/\mu_f(1)$ obtained as the squarefree level q and the newform f vary is equicontinuous for the supremum norm on $C_c(Y_0(1))$, since

$$\left| \frac{\mu_f(\phi_1)}{\mu_f(1)} - \frac{\mu_f(\phi_2)}{\mu_f(1)} \right| \leq \sup |\phi_1 - \phi_2|$$

for any bounded functions ϕ_1, ϕ_2 on $Y_0(1)$. Thus Theorem 1.3 follows if we can show that $\mu_f(\phi)/\mu_f(1) \to \mu(\phi)/\mu(1)$ as $q \to \infty$ for a set of bounded functions ϕ the uniform closure of whose span contains $C_c(Y_0(1))$.

Such a spanning set is furnished [19] by the the Maass eigencuspforms and the incomplete Eisenstein series, as defined in §2. We have $\mu(\phi) = 0$ for a Maass eigencuspform ϕ because ϕ and the constant function 1 lie in orthogonal eigenspaces of the (self-adjoint) hyperbolic Laplacian, so we obtain the following reformulation of our main theorem.

Lemma 1.10 ("Weyl's criterion")**.** *Suppose that for each fixed Maass eigencuspform ϕ, we have*

$$\frac{\mu_f(\phi)}{\mu_f(1)} \to 0 \qquad \textit{as } q \to \infty,$$

and for each fixed incomplete Eisenstein series ϕ, we have

$$\frac{\mu_f(\phi)}{\mu_f(1)} \to \frac{\mu(\phi)}{\mu(1)} \qquad \textit{as } q \to \infty;$$

in neither case need the convergence be uniform in ϕ. Then Theorem 1.3 is true.

1.6. **Acknowledgements.** We thank Dinakar Ramakrishnan for suggesting this problem and for his very helpful feedback and comments on earlier drafts of this paper. We also thank Abhishek Saha for his careful reading of and useful comments on an earlier draft.

2. Background on automorphic forms

We collect here some standard properties of classical automorphic forms. We refer to Serre [35], Shimura [36], Iwaniec [18, 19] and Atkin-Lehner [1] for complete definitions and proofs.

2.1. **Holomorphic newforms.** Let k be a positive even integer, and let α be an element of $\mathrm{GL}(2,\mathbb{R})$ with positive determinant; the element α acts on $\mathbb{H}$ by fractional linear transformations in the usual way. Given a function $f : \mathbb{H} \to \mathbb{C}$, we denote by $f|_k\alpha$ the function $z \mapsto \det(\alpha)^{k/2} j(\alpha, z)^{-k} f(\alpha z)$, where $j\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right), z\right) = cz + d$.

A *holomorphic cusp form* on $\Gamma_0(q)$ of weight k is a holomorphic function $f : \mathbb{H} \to \mathbb{C}$ that satisfies $f|_k\gamma = f$ for all $\gamma \in \Gamma_0(q)$ and vanishes at the cusps of $\Gamma_0(q)$. A *holomorphic newform* is a cusp form that is an eigenform of the algebra of Hecke operators and orthogonal with respect to the Petersson inner product to the oldforms.[3] We say that a holomorphic newform f is a *normalized holomorphic newform* if moreover $\lambda_f(1) = 1$ in the Fourier expansion

$$y^{k/2} f(z) = \sum_{n \in \mathbb{N}} \frac{\lambda_f(n)}{\sqrt{n}} \kappa_f(ny) e(nx), \tag{4}$$

[3] The terms we leave undefined are standard and their precise definitions, which may be found in the references mentioned above, are not necessary for our purposes.

where $\kappa_f(y) = y^{k/2}e^{-2\pi y}$ and $e(x) = e^{2\pi i x}$; in that case the Fourier coefficients $\lambda_f(n)$ are real, multiplicative, and satisfy [4, 5] the Deligne bound $|\lambda_f(n)| \leq \tau(n)$, where $\tau(n)$ denotes the number of positive divisors of n. If $\gamma \in \Gamma_0(q)$ and $z' = \gamma z = x' + iy'$, then $y'^{k/2}f(z') = (j(\gamma,z)/|j(\gamma,z)|)^k y^{k/2} f(z)$, so that in particular $z \mapsto y^k|f(z)|^2$ is $\Gamma_0(q)$-invariant and our definition of μ_f given in Section 1.4 makes sense.

To a newform f one attaches the finite part of the adjoint L-function $L(\operatorname{ad} f, s) = \prod_p L_p(\operatorname{ad} f, s)$ and its completion $\Lambda(\operatorname{ad} f, s) = L_\infty(\operatorname{ad} f, s)L(\operatorname{ad} f, s) = \prod_v L_v(\operatorname{ad} f, s)$, where p traverses the set of primes and v the set of places of $\mathbb{Q}$; the local factors $L_v(\operatorname{ad} f, s)$ are as in [42, §3.1.1]. The Rankin-Selberg method [28, 33] and a standard calculation [42, §3.2.1] show that

$$\mu_f(1) := \int_{\Gamma_0(q)\backslash\mathbb{H}} |f|^2(z) y^k \frac{dx\,dy}{y^2} = q\frac{\Gamma(k-1)}{(4\pi)^{k-1}}\frac{k-1}{2\pi^2}L(\operatorname{ad} f, 1) \asymp_k qL(\operatorname{ad} f, 1). \tag{5}$$

As in the analogous weight aspect [14, p7], the work of Gelbart-Jacquet [8] (following Shimura [37]) and the theorem of Hoffstein-Lockhart [13, Theorem 0.1] (with appendix by Goldfeld-Hoffstein-Lieman) imply that

$$L(\operatorname{ad} f, 1)^{-1} \ll_k \log(eq), \tag{6}$$

where e is Euler's number.

Let σ traverse a set of representatives for the double coset space $\Gamma_\infty\backslash\Gamma/\Gamma_0(q)$. Then the points $\mathfrak{a}_\sigma := \sigma^{-1}\infty \in \mathbb{P}^1(\mathbb{Q})$ traverse a set of inequivalent cusps of $\Gamma_0(q)$. The integer $d_\sigma := [\Gamma_\infty : \Gamma_\infty \cap \sigma\Gamma_0(q)\sigma^{-1}]$ is the width of the cusp $\mathfrak{a}_\sigma$, while

$$w_\sigma := \sigma^{-1}\begin{pmatrix} d_\sigma & \\ & 1 \end{pmatrix}$$

is the scaling matrix for $\mathfrak{a}_\sigma$, which means that $z \mapsto z_\sigma := w_\sigma z$ is a proper isometry of $\mathbb{H}$ under which $z_\sigma \mapsto z_\sigma + 1$ corresponds to the action on z by a generator for the $\Gamma_0(q)$-stabilizer of $\mathfrak{a}_\sigma$.

If the bottom row of σ^{-1} is (c, d), then $d_\sigma = q/(q, c^2)$; moreover, as σ varies, the multiset of widths $\{d_\sigma\}$ is the set $\{d : d|q\}$ of positive divisors of q [19, §2.4]. In particular, c and d_σ are coprime, so we may and shall assume (after multiplying σ on the left by an element of Γ_∞ if necessary) that d_σ divides d. Since q is squarefree, the numbers d_σ and q/d_σ are coprime, so that w_σ is an Atkin-Lehner operator "W_Q" in the sense of [1, p138]. Thus by applying [1, Thm 3] to the newform f, we obtain

$$f|_k w_\sigma = \pm f. \tag{7}$$

Since f is $\Gamma_0(q)$-invariant, the property (7) does not depend upon the choice of coset representatives σ.

2.2. **Maass eigencuspforms.** A *Maass cusp form* (of level 1) is a Γ-invariant eigenfunction of the hyperbolic Laplacian $\Delta := y^{-2}(\partial_x^2 + \partial_y^2)$ on $\mathbb{H}$ that decays rapidly at the cusp of Γ. By Selberg's "$\lambda_1 \geq 1/4$" theorem [34] there exists a real number $r \in \mathbb{R}$ such that $(\Delta + 1/4 + r^2)\phi = 0$; our arguments use only that $r \in \mathbb{R} \cup i(-1/2, 1/2)$, and so apply verbatim in contexts where "$\lambda_1 \geq 1/4$" is not known.

A *Maass eigencuspform* is a Maass cusp form that is an eigenfunction of the (non-archimedean) Hecke operators and the involution $T_{-1} : \phi \mapsto [z \mapsto \phi(-\bar{z})]$, which commute one another as well as with Δ. A Maass eigencuspform ϕ has a Fourier expansion

$$\phi(z) = \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{\lambda_\phi(n)}{\sqrt{|n|}} \kappa_{ir}(ny) e(nx) \tag{8}$$

where $\kappa_{ir}(y) = 2|y|^{1/2}K_{ir}(2\pi|y|)\operatorname{sgn}(y)^{\frac{1+\delta}{2}}$ with K_{ir} the standard K-Bessel function, sgn the signum function, and $\delta \in \{\pm 1\}$ the T_{-1}-eigenvalue of ϕ. We have $|\kappa_s(y)| \leq 1$ for all $s \in i\mathbb{R} \cup (-1/2, 1/2)$ and all $y \in \mathbb{R}_+^*$. A *normalized Maass eigencuspform* further satisfies $\lambda_\phi(1) = 1$; in that case the coefficients $\lambda_\phi(n)$ are real, multiplicative, and satisfy, for each $x \geq 1$, the Rankin-Selberg bound [19, Theorem 3.2]

$$\sum_{n \leq x} |\lambda_\phi(n)|^2 \ll_\phi x. \tag{9}$$

Because $f(-\bar{z}) = \overline{f(z)}$ for any normalized holomorphic eigencuspform f, we have $\mu_f(\phi) = 0$ whenever $T_{-1}\phi = \delta\phi$ with $\delta = -1$. Thus we shall assume throughout this paper that $\delta = 1$, i.e., that ϕ is an *even* Maass form.

2.3. **Eisenstein series.** Let $s \in \mathbb{C}$ and $z \in \mathbb{H}$. The *real-analytic Eisenstein series* $E(s,z) = \sum_{\Gamma_\infty \backslash \Gamma} \operatorname{Im}(\gamma z)^s$ converges normally for $\operatorname{Re}(s) > 1$ and continues meromorphically to the half-plane $\operatorname{Re}(s) \geq 1/2$ where the map $s \mapsto E(s,z)$ is holomorphic with the exception of a unique simple pole at $s = 1$ of constant residue $\operatorname{res}_{s=1} E(s,z) = \mu(1)^{-1}$. The Eisenstein series satisfies the invariance $E(s,\gamma z) = E(s,z)$ for all $\gamma \in \Gamma$ and admits the Fourier expansion

$$E(s,z) = y^s + M(s) y^{1-s} + \frac{1}{\xi(2s)} \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{\lambda_{s-1/2}(n)}{\sqrt{|n|}} \kappa_{s-1/2}(ny) e(nx), \tag{10}$$

where $\lambda_s(n) = \sum_{ab=n} (a/b)^s$, $\kappa_s(y) = 2|y|^{1/2} K_s(2\pi|y|)$, $M(s) = \xi(2s-1)/\xi(2s)$, $\xi(s) = \Gamma_{\mathbb{R}}(s)\zeta(s)$, $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$, and $\zeta(s) = \sum_{n \in \mathbb{N}} n^{-s}$ (for $\operatorname{Re}(s) > 1$) is the Riemann zeta function. The identity $|M(s)| = 1$ for $\operatorname{Re}(s) = 1/2$ follows from (for instance) the functional equation for the zeta function and the prime number theorem. When $\operatorname{Re}(s) = 1/2$ we call $E(s,z)$ a *unitary Eisenstein series.*

2.4. **Incomplete Eisenstein series.** Let $\Psi \in C_c^\infty(\mathbb{R}_+^*)$ be a nonnegative-valued test function with Mellin transform $\Psi^\wedge(s) = \int_0^\infty \Psi(y) y^{-s-1}\, dy$. Repeated partial integration shows that $|\Psi^\wedge(s)| \ll_{\Psi,A} (1+|s|)^A$ for each positive integer A, uniformly for s in vertical strips. The Mellin inversion formula asserts that $\Psi(y) = \int_{(2)} \Psi^\wedge(s) y^s \frac{ds}{2\pi i}$, where $\int_{(\sigma)}$ denotes the integral taken over the vertical contour from $\sigma - i\infty$ to $\sigma + i\infty$. To such Ψ we attach the *incomplete Eisenstein series*

$$E(\Psi, z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \Psi(\operatorname{Im}(\gamma z)). \tag{11}$$

The sum has a uniformly bounded finite number of nonzero terms for z in a fixed compact subset of $\mathbb{H}$. By Mellin inversion, the rapid decay of $\Psi^\wedge$ and Cauchy's theorem, we have

$$E(\Psi, z) = \int_{(2)} \Psi^\wedge(s) E(s,z) \frac{ds}{2\pi i} = \frac{\Psi^\wedge(1)}{\operatorname{vol}(\Gamma \backslash \mathbb{H})} + \int_{(1/2)} \Psi^\wedge(s) E(s,z) \frac{ds}{2\pi i}. \tag{12}$$

Let $\phi = E(\Psi, \cdot)$ be an incomplete Eisenstein series. Note that $\mu(\phi) = \Psi^\wedge(1)$. By comparing (12) and (10), we may express the Fourier coefficients $\phi_n(y)$ in the Fourier series $\phi(z) = \sum_{n \in \mathbb{Z}} \phi_n(y) e(nx)$ as

$$\phi_n(y) = \int_{(1/2)} \frac{\Psi^\wedge(s)}{\xi(2s)} \frac{\lambda_{s-1/2}(n)}{\sqrt{|n|}} \kappa_{s-1/2}(ny) \frac{ds}{2\pi i} \qquad (n \neq 0), \tag{13}$$

$$\phi_0(y) = \frac{\mu(\phi)}{\mu(1)} + \int_{(1/2)} \Psi^\wedge(s) \left(y^s + M(s) y^{1-s}\right) \frac{ds}{2\pi i} \qquad (n = 0). \tag{14}$$

3. Main estimates

We prove a level aspect analogue of Holowinsky's main bound [15, Corollary 3]. To formulate our result, define for each normalized holomorphic newform f and each real number $x \geq 1$ the quantities

$$M_f(x) = \frac{\prod_{p \leq x} (1 + 2|\lambda_f(p)|/p)}{\log(ex)^2 L(\operatorname{ad} f, 1)}, \quad R_f(x) = \frac{x^{-1/2}}{L(\operatorname{ad} f, 1)} \int_{\mathbb{R}} \left| \frac{L(\operatorname{ad} f, \frac{1}{2} + it)}{(1+|t|)^{10}} \right| dt. \tag{15}$$

In §5 we shall refer only to the definitions (15) and the statement of the following theorem, not its proof.

Theorem 3.1. *Let f be a normalized holomorphic newform of weight k and squarefree level q. If ϕ is a Maass eigencuspform, then*

$$\frac{\mu_f(\phi)}{\mu_f(1)} \ll_{k,\phi,\varepsilon} \log(eq)^{\varepsilon} M_f(q)^{1/2}.$$

If ϕ is an incomplete Eisenstein series, then

$$\frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} \ll_{k,\phi,\varepsilon} \log(eq)^{\varepsilon} M_f(q)^{1/2} \left(1 + R_f(q)\right).$$

In this section k is a positive even integer, f is a normalized holomorphic newform of weight k and squarefree level q, and ϕ is a Maass eigencuspform or incomplete Eisenstein series. In §3.1 we reduce Theorem 3.1 to a problem of estimating shifted sums (see Definition 3.2). In §3.2 we apply a refinement of [15, Theorem 2] to bound such shifted sums. In §3.3 we complete the proof of Theorem 3.1.

3.1. **Reduction to shifted sums.** Fix once and for all an everywhere nonnegative test function $h \in C_c^{\infty}(\mathbb{R}_+^*)$ with Mellin transform $h^\wedge(s) = \int_0^\infty h(y) y^{-s-1}\, dy$ such that $h^\wedge(1) = \mu(1)$. In what follows, all implied constants may depend upon h without mention.

Definition 3.2. To the parameters $s \in \mathbb{C}$, $l \in \mathbb{Z}_{\neq 0}$ and $x \geq 1$ we associate the *shifted sums*

$$S_s(l,x) = \sum_{\substack{n \in \mathbb{N} \\ m := n+l \in \mathbb{N}}} \frac{\lambda_f(m)}{\sqrt{m}} \frac{\lambda_f(n)}{\sqrt{n}} I_s(l,n,x),$$

where $I_s(l,n,x)$ is an integral depending upon our fixed test function h:

$$I_s(l,n,x) = \int_0^\infty h(xy) \kappa_s(ly) \kappa_f(my) \kappa_f(ny) y^{-1} \frac{dy}{y}, \quad m := n + l.$$

Our aim in this section is to reduce Theorem 3.1 to the problem of bounding such shifted sums. We shall subsequently refer to the statement below of Proposition 3.3 but not the details of its proof.

Proposition 3.3. *Let $Y \geq 1$. If ϕ is a Maass eigencuspform of eigenvalue $1/4 + r^2$, then*

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{1}{Y \mu_f(1)} \sum_{\substack{l \in \mathbb{Z}_{\neq 0} \\ |l| < Y^{1+\varepsilon}}} \frac{\lambda_\phi(l)}{\sqrt{|l|}} \sum_{d|q} S_{ir}(dl, dY) + O_{\phi,\varepsilon}(Y^{-1/2}).$$

If $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series, then

$$\frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} = \frac{1}{Y \mu_f(1)} \int_{\mathbb{R}} \frac{\Psi^\wedge(\frac{1}{2} + it)}{\xi(1 + 2it)} \left(\sum_{\substack{l \in \mathbb{Z}_{\neq 0} \\ |l| < Y^{1+\varepsilon}}} \frac{\lambda_{it}(l)}{\sqrt{|l|}} \sum_{d|q} S_{it}(dl, dY) \right) \frac{dt}{2\pi}$$
$$+ O_{\phi,\varepsilon}\left(\frac{1 + R_f(q)}{Y^{1/2}} \right).$$

Our proof follows a sequence of lemmas. Let k, f, q, Y, ϕ, h be as above and let h_Y be the function $y \mapsto h(Yy)$. To h_Y we attach the incomplete Eisenstein series $E(h_Y, z)$ by the usual recipe (11).

Lemma 3.4. *We have the following approximate formula for the quantity $\mu_f(\phi)$:*

$$Y \mu_f(\phi) = \sum_{d|q} \int_{y=0}^\infty h_Y(dy) \int_{x=0}^1 \phi(dy) |f|^2(z) y^k \frac{dx\, dy}{y^2} + O_\phi(Y^{1/2} \mu_f(1)).$$

Proof. By Mellin inversion and Cauchy's theorem as in (12), we have

$$Y\mu_f(\phi) = \mu_f(E(h_Y,\cdot)\phi) - \int_{(1/2)} h^\wedge(s)Y^s\mu_f(E(s,\cdot)\phi)\,\frac{ds}{2\pi i}.$$

The argument of [15, Proof of Lemma 3.1a] shows without modification that

$$\int_{(1/2)} h^\wedge(s)Y^s\mu_f(E(s,\cdot)\phi)\,\frac{ds}{2\pi i} \ll_\phi Y^{1/2}\mu_f(1); \tag{16}$$

since the proof is short, we sketch it here. By the Fourier expansion for $E(s,z)$ and the rapid decay of $\phi(z)$ as $y\to\infty$, we have $E(s,z)\phi(z) \ll_\phi |s|^{O(1)}$ for $\mathrm{Re}(s)=1/2$ and z in the Siegel domain $\{z : x\in[0,1], y>1/2\}$ for $\Gamma\backslash\mathbb{H}$. By the rapid decay of $h^\wedge$ we have $h^\wedge(s)Y^sE(s,z)\phi(z) \ll_\phi Y^{1/2}|s|^{-2}$ for s,z as above; the estimate (16) follows by integrating in z against μ_f and then integrating in s.

Having established that $Y\mu_f(\phi) = \mu_f(E(h_Y,\cdot)\phi) + O_\phi(Y^{1/2}\mu_f(1))$, it remains now only to evaluate $\mu_f(E(h_Y,\cdot)\phi)$. Let $\Gamma_\infty\backslash\Gamma/\Gamma_0(q) = \{\sigma\}$ be a set of double-coset representatives as in §2.1, and set

$$d_\sigma = [\Gamma_\infty : \Gamma_\infty\cap\sigma\Gamma_0(q)\sigma^{-1}].$$

By decomposing the transitive right Γ-set $\Gamma_\infty\backslash\Gamma$ into $\Gamma_0(q)$-orbits

$$\Gamma_\infty\backslash\Gamma = \sqcup\Gamma_\infty\backslash\Gamma_\infty\sigma\Gamma_0(q) = \sqcup\sigma(\sigma^{-1}\Gamma_\infty\sigma\cap\Gamma_0(q)\backslash\Gamma_0(q)),$$

we obtain

$$E(h_Y,z) = \sum_{\substack{\sigma\in\Gamma_\infty\backslash\Gamma/\Gamma_0(q)\\ \gamma\in\sigma^{-1}\Gamma_\infty\sigma\cap\Gamma_0(q)\backslash\Gamma_0(q)}}\sum h_Y(\mathrm{Im}(\sigma\gamma z)).$$

By invoking the $\Gamma_0(q)$-invariance of $z\mapsto\phi(z)|f|^2(z)y^k\,\frac{dx\,dy}{y^2}$ and unfolding the sum over $\gamma\in\sigma^{-1}\Gamma_\infty\sigma\cap\Gamma_0(q)\backslash\Gamma_0(q)$ with the integral over $z\in\Gamma_0(q)\backslash\mathbb{H}$, we get

$$\mu_f(E(h_Y,\cdot)\phi) = \sum_{\sigma\in\Gamma_\infty\backslash\Gamma/\Gamma_0(q)}\int_{\sigma^{-1}\Gamma_\infty\sigma\cap\Gamma_0(q)\backslash\mathbb{H}} h_Y(\mathrm{Im}(\sigma z))\phi(z)|f|^2(z)y^k\,\frac{dx\,dy}{y^2}.$$

The change of variables $z\mapsto\sigma^{-1}z$ transforms the integral above into

$$\int_{\Gamma_\infty\cap\sigma\Gamma_0(q)\sigma^{-1}\backslash\mathbb{H}} h_Y(y)\phi(z)|f|^2(\sigma^{-1}z)\mathrm{Im}(\sigma^{-1}z)^k\,\frac{dx\,dy}{y^2}.$$

Integrating over a fundamental domain for $\Gamma_\infty\cap\sigma\Gamma_0(q)\sigma^{-1} = \{\pm\left(\begin{smallmatrix}1 & d_\sigma n\\ & 1\end{smallmatrix}\right) : n\in\mathbb{Z}\}$ acting on $\mathbb{H}$, we get

$$\int_{y=0}^{\infty} h_Y(y)\int_{x=0}^{d_\sigma}\phi(z)|f|^2(\sigma^{-1}z)\mathrm{Im}(\sigma^{-1}z)^k\,\frac{dx\,dy}{y^2}.$$

Applying now the change of variables $z\mapsto d_\sigma z$ gives

$$\int_{y=0}^{\infty} h_Y(d_\sigma y)\int_{x=0}^{1}\phi(d_\sigma z)\left|f|_k\sigma^{-1}\left(\begin{smallmatrix}d_\sigma & \\ & 1\end{smallmatrix}\right)\right|^2(z)y^k\,\frac{dx\,dy}{y^2}.$$

Since $f|_k\sigma^{-1}\left(\begin{smallmatrix}d_\sigma & \\ & 1\end{smallmatrix}\right) = \pm f$ by the consequence (7) of Atkin-Lehner theory (using here that q is squarefree), we find that

$$\mu_f(E(h_Y,\cdot)\phi) = \sum_{\sigma\in\Gamma_\infty\backslash\Gamma/\Gamma_0(q)}\int_{y=0}^{\infty} h_Y(d_\sigma y)\int_{x=0}^{1}\phi(d_\sigma z)|f|^2(z)y^k\,\frac{dx\,dy}{y^2}.$$

Since $\{d_\sigma\} = \{d : d|q\}$, we obtain the claimed formula. □

In the expression for $Y\mu_f(\phi)$ given by Lemma 3.4, we expand ϕ in a Fourier series $\phi(z) = \sum_{l\in\mathbb{Z}} \phi_l(y)e(lx)$ and consider separately the contributions from l in various ranges; specifically, we set

$$\mathcal{S}_0 = \sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} \phi_0(dy)|f|^2(z)y^k \,\frac{dx\,dy}{y^2},$$

$$\mathcal{S}_{(0,Y^{1+\varepsilon})} = \sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} \sum_{0<|l|<Y^{1+\varepsilon}} \phi_l(dy)|f|^2(z)y^k \,\frac{dx\,dy}{y^2},$$

$$\mathcal{S}_{\geq Y^{1+\varepsilon}} = \sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} \sum_{|l|\geq Y^{1+\varepsilon}} \phi_l(dy)|f|^2(z)y^k \,\frac{dx\,dy}{y^2},$$

so that

$$\sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} \phi(dz)|f|^2(z)y^k \,\frac{dx\,dy}{y^2} = \mathcal{S}_0 + \mathcal{S}_{(0,Y^{1+\varepsilon})} + \mathcal{S}_{\geq Y^{1+\varepsilon}}. \tag{17}$$

We treat these contributions in Lemmas 3.6, 3.7 and 3.8, respectively; in doing so we shall repeatedly use the following technical result.

Lemma 3.5. *The quantity $\mu_f(E(h_Y,\cdot))$ satisfies the formulas and estimates*

$$\begin{aligned}\mu_f(E(h_Y,\cdot)) &= \sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} |f|^2(z)y^k \,\frac{dx\,dy}{y^2} \\ &= Y\mu_f(1)\,(1+E_f(qY)) \\ &= Y\mu_f(1)\left(1+O_k\left(Y^{-1/2}R_f(q)\right)\right),\end{aligned}$$

where

$$E_f(x) := \frac{2\pi^2}{x}\int_{(1/2)} h^\wedge(s)\left(\frac{x}{4\pi}\right)^s \frac{\Gamma(s+k-1)}{\Gamma(k)}\,\frac{\zeta(s)}{\zeta(2s)}\,\frac{L(\operatorname{ad} f,s)}{L(\operatorname{ad} f,1)}\,\frac{ds}{2\pi i}.$$

Moreover, $\mu_f(E(h_Y,\cdot)) \ll Y\mu_f(1)$.

Proof. The first equality follows from the same argument as in the proof of Lemma 3.4, the second from the Mellin formula and the unfolding method by a direct computation, the third from the bounds $|\Gamma(k-1/2+it)| \leq \Gamma(k-1/2)|$, $\zeta(1/2+it) \ll (1+|t|)^{1/4}$ and $|\zeta(1+2it)| \gg 1/\log(1+|t|)$ as in [38, p7]. Finally, because the quantity $\mu_f(E(h_Y,\cdot))$ is majorized by the integral of the Γ-invariant measure μ_f over the region on which the function $\Gamma_\infty\backslash\mathbb{H} \ni z \mapsto h_Y(y)$ does not vanish and because that region contains $\ll Y$ fundamental domains for $\Gamma\backslash\mathbb{H}$ [19, Lemma 2.10], we have $\mu_f(E(h_Y,\cdot)) \ll Y\mu_f(1)$. □

Lemma 3.6 (The main term $\mathcal{S}_0$)**.** *If ϕ is a Maass eigencuspform, then $\phi_0(y) = 0$ and $\mathcal{S}_0 = 0$. If ϕ is an incomplete Eisenstein series, then*

$$\mathcal{S}_0 = Y\mu_f(1)\left(\frac{\mu(\phi)}{\mu(1)} + O_\phi\left(\frac{1+R_f(q)}{Y^{1/2}}\right)\right).$$

Proof. If ϕ is a Maass eigencuspform then $\phi_0(y) = 0$ holds by definition, hence $\mathcal{S}_0 = 0$. Suppose otherwise that ϕ is an incomplete Eisenstein series. It follows from (14) that for every $y \in \mathbb{R}_+^*$ such that $h_Y(y) \neq 0$, we have $\phi_0(y) = \mu(\phi)/\mu(1) + O_\phi(Y^{-1/2})$. Thus two applications of Lemma 3.5 show that

$$\begin{aligned}\mathcal{S}_0 &= \mu_f(E(h_Y,\cdot))\left(\frac{\mu(\phi)}{\mu(1)} + O_\phi(Y^{-1/2})\right)\\ &= Y\mu_f(1)\left(1 + O\left(\frac{R_f(q)}{Y^{1/2}}\right)\right)\left(\frac{\mu(\phi)}{\mu(1)} + O_\phi(Y^{-1/2})\right)\\ &= Y\mu_f(1)\left(\frac{\mu(\phi)}{\mu(1)} + O_\phi\left(\frac{1+R_f(q)}{Y^{1/2}}\right)\right).\end{aligned}$$

□

Lemma 3.7 (The essential error term $\mathcal{S}_{(0,Y^{1+\varepsilon})}$). *If ϕ is a Maass eigencuspform, then*

$$\mathcal{S}_{(0,Y^{1+\varepsilon})} = \sum_{0<|l|<Y^{1+\varepsilon}} \frac{\lambda_\phi(l)}{\sqrt{|l|}} \sum_{d|q} S_{ir}(dl, dY).$$

If ϕ is an incomplete Eisenstein series, then

$$\mathcal{S}_{(0,Y^{1+\varepsilon})} = \int_{\mathbb{R}} \frac{\Psi^\wedge(\frac{1}{2}+it)}{\xi(1+2it)} \sum_{0<|l|<Y^{1+\varepsilon}} \frac{\lambda_{it}(l)}{\sqrt{|l|}} \sum_{d|q} S_{it}(dl, dY)\,\frac{dt}{2\pi}$$

Proof. Follows by integrating the Fourier expansion (4) of a newform, the Fourier expansion (8) of a Maass cusp form, and the formula (13) for the non-constant Fourier coefficients of an Eisenstein series. □

Lemma 3.8 (The trivial error term $\mathcal{S}_{\geq Y^{1+\varepsilon}}$). *We have $\mathcal{S}_{\geq Y^{1+\varepsilon}} \ll_{\phi,\varepsilon} Y^{-10}\mu_f(1)$.*

Proof. Lemma 3.8 follows from Lemma 3.5 and the following claim: for all $y \in \mathbb{R}_+^*$ such that $h_Y(y) \neq 0$, we have $\sum_{|l|\geq Y^{1+\varepsilon}} |\phi_l(y)| \ll_{\phi,\varepsilon} Y^{-11}$. The claim is proved in [15, §3.2], as follows. When ϕ is a cusp form of eigenvalue $1/4 + r^2$, so that $\phi_l(y) = y^{-1/2}\lambda_\phi(l)\kappa_{ir}(ly)$, the claim follows from the exponential decay of $l \mapsto \kappa_{ir}(ly)$ for $l \geq Y^{1+\varepsilon}$ and $y \asymp Y^{-1}$ together with the polynomial growth of $l \mapsto \lambda_\phi(l)$. When ϕ is an incomplete Eisenstein series, the integral formula (13) and standard bounds for the K-Bessel function show that for each positive integer A, we have $\phi_l(y) \ll_{\phi,\varepsilon,A} \tau(l)Y^{A-1/2}|l|^{-A}(1 + Y/|l|)^\varepsilon$; the claim then follows by summing over $|l| \geq Y^{1+\varepsilon}$. □

Proof of Proposition 3.3. By Lemma 3.4 and equation (17), we have

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{1}{Y\mu_f(1)}\left(\mathcal{S}_0 + \mathcal{S}_{(0,Y^{1+\varepsilon})} + \mathcal{S}_{\geq Y^{1+\varepsilon}}\right) + O_{\phi,\varepsilon}(Y^{-1/2}).$$

Proposition 3.3 follows by combining the results of Lemma 3.6, Lemma 3.8 and Lemma 3.7. □

3.2. **Bounds for individual shifted sums.** We bound the individual shifted sums appearing in Definition 3.2; in subsequent sections we shall need only our main result, Corollary 3.14. We first recall a special case of Holowinsky's bound [15, Theorem 2].

Theorem 3.9 (Holowinsky). *Let $\varepsilon \in (0,1)$. Then for $x \geq 1$ and $l \in \mathbb{Z}_{\neq 0}$, we have*

$$\sum_{\substack{n\in\mathbb{N}\\ m:=n+l\in\mathbb{N}\\ \max(m,n)\leq x}} |\lambda_f(m)\lambda_f(n)| \ll_\varepsilon \tau(l)\frac{x\prod_{p\leq x}(1+2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}}$$

Unfortunately, Theorem 3.9 is insufficient for our purposes because $\tau(ql)$ can be quite large, even larger asymptotically than every power of $\log(eq)$, when q has many small prime factors. The following refinement will suffice.

Theorem 3.10. *With conditions as in the statement of Theorem 3.9, we have*

$$\sum_{\substack{n\in\mathbb{N}\\ m:=n+l\in\mathbb{N}\\ \max(m,n)\leq x}} |\lambda_f(m)\lambda_f(n)| \ll_\varepsilon \frac{x\prod_{p\leq x}(1+2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}} \tag{18}$$

Proof. In [26, Thm 3.1], we generalized Holowinsky's bound [15, Thm 2] to totally real number fields. Along the way we proved a pair of results [26, Thm 4.10] and [26, Thm 7.2] either of which give Theorem 3.10 in the special case $\mathbb{F}=\mathbb{Q}$. We could just as easily give the proof here in the special case $\mathbb{F}=\mathbb{Q}$, but spare ourselves the trouble since we have already carried it out in greater generality; let us also emphasize that our proof is self-contained and does not depend upon other literature concerning QUE and its generalizations. ☐

Remark 3.11. A bound of the form (18) but with an *unspecified* dependence on the parameter l may be derived from the work of Nair [25]. We have attempted to quantify this dependence by working through the details of Nair's arguments, and have shown that they imply

$$\sum_{\substack{n\in\mathbb{N}\\ m:=n+l\in\mathbb{N}\\ \max(m,n)\leq x}} |\lambda_f(m)\lambda_f(n)| \ll_\varepsilon \tau_m(l)\frac{x\prod_{p\leq x}(1+2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}} \tag{19}$$

for some $m\geq 2$ (probably $m=2$) and all $0\neq |l|\leq x^{1/16-\varepsilon}$; in deducing this we have used the Ramanujan bound $|\lambda_f(p)|\leq 2$. This strength and uniformity falls far short of what is needed in treating the level aspect of QUE.

A mild strengthening of (18) subject to the additional constraint $4l^2\leq x$ appears in the recent book of Iwaniec-Friendlander [6, Thm 15.6], which was released after we completed the work of this paper. The condition $4l^2\leq x$ makes their result inapplicable in our treatment of the level aspect of QUE, where l can be nearly as large as x. However, it seems to us that one can remove this condition by a suitable modification of their arguments.

Recall from Definition 3.2 that the sums $S_s(l,x)$ involve a certain integral $I_s(l,n,x)$.

Lemma 3.12. *For each positive integer A, the integral $I_s(l,n,x)$ satisfies the upper bound*

$$I_s(l,n,x)\ll_{k,A}\sqrt{mn}\cdot\min\left(1,\frac{\max(m,n)}{x}\right)^{-A}$$

uniformly for $s\in i\mathbb{R}\cup(-1/2,1/2)$, $n\in\mathbb{N}$, $l\in\mathbb{Z}_{\neq 0}$, and $x\geq 1$. Here $m:=n+l$, as usual.

Proof. Let s,l,m,n be as above, and let $A\geq 0$. Then $|\kappa_s(y)|\leq 1$, so that by the Mellin formula we have

$$\begin{aligned}
I_s(l,n,x) &\leq \int_0^\infty h(xy)\kappa_f(my)\kappa_f(ny)y^{-1}\,\frac{dy}{y}\\
&= \int_{(A)} h^\wedge(w)x^w\int_{\mathbb{R}_+^*} y^{w-1}\kappa_f(my)\kappa_f(ny)\,\frac{dy}{y}\,\frac{dw}{2\pi i}\\
&= \frac{(\sqrt{mn})^k}{\left(4\pi\left(\frac{m+n}{2}\right)\right)^{k-1}}\int_{(A)} h^\wedge(w)\left(\frac{x}{4\pi\left(\frac{m+n}{2}\right)}\right)^w\Gamma(w+k-1)\,\frac{dw}{2\pi i}\\
&\ll_{k,A} \sqrt{mn}\left(\frac{\max(m,n)}{x}\right)^{-A}.
\end{aligned}$$

Here we used the arithmetic mean-geometric mean inequality, the elementary bound $\Gamma(w+k-1)\ll_{k,A}1$ for $\mathrm{Re}(w)=A$, and the rapid decay of $h^\wedge$. The case $A=0$ gives $I_s(l,n,x)\ll_k\sqrt{mn}$, which combined with the case that A is a positive integer yields the assertion of the lemma. ☐

Remark 3.13. See [26, Lem 4.3] and [26, Cor 4.4] for a fairly sharp refinement of Lemma 3.12.

Corollary 3.14. *The shifted sums $S_s(l,x)$ satisfy the upper bound*

$$S_s(l,x) \ll_{k,\varepsilon} \frac{x\prod_{p\leq x}(1+2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}} \tag{20}$$

uniformly for $s \in i\mathbb{R} \cup (-1/2,1/2)$ and $x \geq 1$.

Proof. Let us temporarily denote by $T_f(x,l,\varepsilon)$ the right hand side of (20). By Definition 3.2 and Lemma 3.12, we need only show that

$$\sum_{\substack{n\in\mathbb{N}\\ m:=n+l\in\mathbb{N}}} |\lambda_f(m)\lambda_f(n)| \cdot \min\left(1, \frac{\max(m,n)}{x}\right)^{-A} \ll_\varepsilon T_f(x,l,\varepsilon) \tag{21}$$

for some positive integer A. Take $A = 2$. We may assume that $x \geq 10$. By Theorem 3.10 and the Deligne bound $|\lambda_f(p)| \leq 2$, the left hand side of (21) is

$$\begin{aligned} &\ll_\varepsilon \quad T_f(x,l,\varepsilon)\sum_{n=0}^{\infty} 2^{-nA}2^n\left(\frac{\log(ex)}{\log(e2^nx)}\right)^{2-\varepsilon}\prod_{x<p\leq 2^nx}(1+2|\lambda_f(p)|/p)\\ &\ll \quad T_f(x,l,\varepsilon)\sum_{n=0}^{\infty} 2^{-(A-1)n}\exp\left(4\log\frac{\log(2^nx)}{\log(x)}\right).\end{aligned}$$

The inner sum converges and is bounded uniformly in x, so we obtain the desired estimate (21). □

3.3. **Bounds for sums of shifted sums.** We complete the proof of Theorem 3.1 by bounding the sums of shifted sums that arose in Proposition 3.3.

Lemma 3.15. *For each $\varepsilon \in (0,1)$ and each squarefree number q, we have*

$$\sum_{d|q}\frac{d}{\log(ed)^{2-\varepsilon}} \ll \frac{q\log\log(e^2q)}{\log(eq)^{2-\varepsilon}} \ll_\varepsilon \frac{q}{\log(eq)^{2-2\varepsilon}}.$$

Proof. Suppose that q is the product of $r \geq 1$ primes $q_1 < \cdots < q_r$. Let $p_1 < \cdots < p_r$ be the first r primes, so that $p_i \leq q_i$ for $i = 1,\dots,r$. Define $\beta(x) = x/\log(e^2x)^{2-\varepsilon}$; we have chosen this particular definition so that β is increasing on $\mathbb{R}_{\geq 1}$ and $\beta(x) \asymp x/\log(ex)^{2-\varepsilon}$ for $x \in \mathbb{R}_{\geq 1}$. The map

$$\mathbb{R}_{\geq 1} \ni x \mapsto \log\beta(e^x) = x - (2-\varepsilon)\log(2+x)$$

is convex, so that for each $a = (a_1,\dots,a_r) \in \{0,1\}^r$ we have

$$\frac{\beta(q_1^{a_1}\cdots q_r^{a_r})}{\beta(q_1\cdots q_r)} \leq \frac{\beta(p_1^{a_1}q_2^{a_2}\cdots q_r^{a_r})}{\beta(p_1q_2\cdots q_r)} \leq \frac{\beta(p_1^{a_1}p_2^{a_2}q_3^{a_3}\cdots q_r^{a_r})}{\beta(p_1p_2q_3\cdots q_r)} \leq \cdots \leq \frac{\beta(p_1^{a_1}\cdots p_r^{a_r})}{\beta(p_1\cdots p_r)}. \tag{22}$$

The prime number theorem implies that $\log(p_1\cdots p_r) = r\log(r)(1+o(1))$, where the notation $o(1)$ refers to the limit as $r \to \infty$; we may and shall assume that r is sufficiently large (and at least 100) because the assertion of the lemma holds trivially when q has a bounded number of prime factors. Set $r_0 = \lfloor r/10 \rfloor$. Observe that

$$\begin{aligned} p_{r-r_0+1}\cdots p_r &= \exp\big(r\log(r) - (r-r_0)\log(r-r_0) + o(r\log(r))\big)\\ &= \exp\left(r_0\log(r) + (r-r_0)\log\left(\frac{r}{r-r_0}\right) + o(r\log(r))\right)\\ &= \exp\left(r_0\log(r)(1+o(1))\right)\\ &\ll (p_1\cdots p_r)^{1/9}\end{aligned} \tag{23}$$

and

$$\log(p_1 \cdots p_{r_0}) = r_0 \log(r_0)(1+o(1)) \asymp r \log(r)(1+o(1)) = \log(p_1 \cdots p_r). \tag{24}$$

Let Ω_0 denote the set of all $a \in \{0,1\}^r$ for which $a_1 + \cdots + a_r \leq r_0$ and Ω_1 the set of all $a \in \{0,1\}^r$ for which $a_1 + \cdots + a_r > r_0$, so that $\{0,1\}^r = \Omega_0 \sqcup \Omega_1$. Then by (23) we have

$$\sum_{a \in \Omega_0} \frac{\beta(p_1^{a_1} \cdots p_r^{a_r})}{\beta(p_1 \cdots p_r)} \leq 2^r \frac{\beta(p_{r-r_0+1} \cdots p_r)}{\beta(p_1 \cdots p_r)} \ll 2^r (p_1 \cdots p_r)^{-7/8} \leq \sqrt[8]{2}. \tag{25}$$

If $a \in \Omega_1$, then (24) implies $\beta(p_1^{a_1} \cdots p_r^{a_r})/\beta(p_1 \cdots p_r) \asymp p_1^{a_1-1} \cdots p_r^{a_r-1}$, so that

$$\sum_{a \in \Omega_1} \frac{\beta(p_1^{a_1} \cdots p_r^{a_r})}{\beta(p_1 \cdots p_r)} \ll \sum_{d | p_1 \cdots p_r} \frac{1}{d} \leq (1+o(1)) e^{\gamma} \log\log(p_1 \cdots p_r) \ll \log\log(e^2 q). \tag{26}$$

Since $\beta(x) \asymp x/\log(ex)^{2-\varepsilon}$ for $x \in \mathbb{R}_{\geq 1}$, it follows from (22), (25) and (26) that

$$\frac{\sum_{d|q} \frac{d}{\log(ed)^{2-\varepsilon}}}{\frac{q}{\log(eq)^{2-\varepsilon}}} \asymp \sum_{d|q} \frac{\beta(d)}{\beta(q)} = \sum_{a \in \{0,1\}^r} \frac{\beta(q_1^{a_1} \cdots q_r^{a_r})}{\beta(q_1 \cdots q_r)} \ll \log\log(e^2 q),$$

which establishes the lemma. □

Corollary 3.16. *Let $Y \geq 1$ with $Y \leq c_1 \log(eq)^{c_2}$ for some $c_1, c_2 \geq 1$. Then our sum of shifted sums satisfies the estimate*

$$\sum_{d|q} S_s(dl, dY) \ll_{k,\varepsilon,c_1,c_2} \frac{qY \prod_{p \leq q}(1 + 2|\lambda_f(p)|/p)}{\log(eq)^{2-\varepsilon}},$$

uniformly for $s \in i\mathbb{R} \cup (-1/2, 1/2)$ and $x \geq 1$.

Proof. By Corollary 3.14, we have

$$\sum_{d|q} S_s(dl, dY) \ll_{k,\varepsilon} Y \left(\prod_{p \leq qY} \left(1 + 2\frac{|\lambda_f(p)|}{p}\right)\right) \sum_{d|q} \frac{d}{\log(ed)^{2-\varepsilon}}. \tag{27}$$

By the Deligne bound $|\lambda_f(p)| \leq 2$, the part of the product in (27) taken over $q < p \leq qY$ is $\ll \log(eY)^4 \ll_{c_1,c_2} \log\log(e^e q)^4$. The claim now follows from Lemma 3.15. □

Lemma 3.17. *Let $\varepsilon > 0$, $Y \geq 1$. If ϕ is a normalized Maass eigencuspform, then*

$$\sum_{0 < |l| < Y^{1+\varepsilon}} \frac{|\lambda_\phi(l)|}{\sqrt{|l|}} \ll_{\phi,\varepsilon} Y^{1/2+2\varepsilon},$$

where (as indicated) the implied constant may depend upon ϕ. On the other hand, if $t \in \mathbb{R}$, then

$$\sum_{0 < |l| < Y^{1+\varepsilon}} \frac{|\lambda_{it}(l)|}{\sqrt{|l|}} \ll_{\varepsilon} Y^{1/2+2\varepsilon},$$

where the implied constant does not depend upon t.

Proof. Follows from the Cauchy-Schwarz inequality, partial summation, the Rankin-Selberg bound (9) for λ_ϕ and the uniform bound $|\lambda_{it}(l)| \leq \tau(l)$ for λ_{it}. □

Proof of Theorem 3.1. Suppose that ϕ is a normalized Maass eigencuspform of eigenvalue $\frac{1}{4}+r^2$. By Proposition 3.3, we have

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{1}{Y\mu_f(1)} \sum_{0<|l|<Y^{1+\varepsilon}} \frac{\lambda_\phi(l)}{\sqrt{|l|}} \sum_{d|q} S_{ir}(dl, dY) + O_{\phi,\varepsilon}(Y^{-1/2}). \tag{28}$$

Recall the formula (5) for $\mu_f(1) \asymp_k qL(\operatorname{ad} f, 1)$ and the definition (15) of $M_f(q)$. We shall ultimately choose $Y \ll_k \log(eq)^{O(1)}$, so Corollary 3.16 gives the bound

$$\frac{1}{Y\mu_f(1)} \sum_{d|q} S_{ir}(dl, dY) \ll_{k,\varepsilon} \log(eq)^\varepsilon M_f(q). \tag{29}$$

By (29) and Lemma 3.17 applied to (28), we find that

$$\begin{aligned} \frac{\mu_f(\phi)}{\mu_f(1)} &\ll_{k,\phi,\varepsilon} \log(eq)^\varepsilon M_f(q) \sum_{0<|l|<Y^{1+\varepsilon}} \frac{|\lambda_\phi(l)|}{\sqrt{|l|}} + Y^{-1/2} \\ &\ll_{\phi,\varepsilon} Y^{1/2+2\varepsilon} \log(eq)^\varepsilon M_f(q) + Y^{-1/2}. \end{aligned}$$

Choosing $Y = \max(1, M_f(q)^{-1}) \ll_k \log(eq)^{O(1)}$ gives the cuspidal case of the theorem.

Suppose now that $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series. Proposition 3.3, Corollary 3.16 and Lemma 3.17 show, as in the cuspidal case, that

$$\begin{aligned} \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} &\ll_{k,\phi,\varepsilon} Y^{1/2+2\varepsilon} \log(eq)^\varepsilon M_f(q) \int_{\mathbb{R}} \left| \frac{\Psi^\wedge(\frac{1}{2}+it)}{\xi(1+2it)} \right| dt + \frac{1+R_f(q)}{Y^{1/2}} \\ &\ll_{k,\phi} Y^{1/2+2\varepsilon} \log(eq)^\varepsilon M_f(q) + \frac{1+R_f(q)}{Y^{1/2}}. \end{aligned}$$

The same choice of Y as above completes the proof. □

4. An extension of Watson's formula

Watson [42], building on earlier work of Garrett [7], Piatetski-Shapiro and Rallis [27], Harris and Kudla [12], and Gross and Kudla [11], proved a beautiful formula relating the integral of the product of three modular forms to the central value of their triple product L-function. Unfortunately, Watson's formula applies only to triples of newforms having the *same* squarefree level. In §5 we shall refer only to the statement of the following extension of Watson's formula to the case of interest, not the details of its proof.

Theorem 4.1. *Let ϕ be a Maass eigencuspform of level 1 and f a holomorphic newform of squarefree level q, as in §2. Then*

$$\frac{\left|\int_{\Gamma_0(q)\backslash\mathbb{H}} \phi(z)|f|^2(z) y^k \frac{dx\,dy}{y^2}\right|^2}{\int_{\Gamma\backslash\mathbb{H}} |\phi|^2(z) y^k \frac{dx\,dy}{y^2} \left(\int_{\Gamma_0(q)\backslash\mathbb{H}} |f|^2(z) y^k \frac{dx\,dy}{y^2}\right)^2} = \frac{1}{8q} \frac{\Lambda(\phi\times f\times f, \frac{1}{2})}{\Lambda(\operatorname{ad}\phi, 1)\Lambda(\operatorname{ad} f, 1)^2}.$$

The L-functions $L(\cdots) = \prod_p L_p(\cdots)$ and their completions $\Lambda(\cdots) = L_\infty(\cdots)L(\cdots) = \prod_v L_v(\cdots)$ are as in [42, §3].

Remark 4.2. For simplicity, we have stated Theorem 4.1 only in the special case that we need it, but our calculations (Lemma 4.4) lead to a slightly more general formula. Let ψ_j $(j = 1, 2, 3)$ be newforms of weight k_j and level q_j. We allow the possibility $k_j = 0$, in which case we require that ψ_j be a Maass form. Then one can read off from Ichino [16] and Lemma 4.4 a formula for the integral of $\psi_1\psi_2\psi_3$ provided that $k_1+k_2+k_3 = 0$, that the q_j are squarefree, that for each prime divisor p of $q_1q_2q_3$ we have $\{v_p(q_1), v_p(q_2), v_p(q_3)\} = \{1, 1, 0\}$ as unordered sets (here $v_p(n)$ is the power to which p divides an integer n).

Watson proved his formula only for three forms of the same squarefree level because Gross and Kudla [11] evaluated the p-adic zeta integrals of Harris and Kudla [12] only when (the factorizable automorphic representations generated by) the three forms are special at p; Harris and Kudla had already considered the case that all three forms are spherical at p. Ichino [16] showed that the local zeta integrals of Harris and Kudla are equal to simpler integrals over the group $\mathrm{PGL}(2, \mathbb{Q}_p)$. Ichino and Ikeda [17, §7, §12] computed these simpler integrals when all three forms are special at p. Since we are interested in the integral of $\phi|f|^2$ when ϕ has level 1 and f has squarefree level q, we must consider the case that two representations are special and one is spherical. We remark in passing that Böcherer and Schulze-Pillot [2] considered similar problems for modular forms on definite rational quaternion algebras in the classical language, but their results are not directly applicable here.

To state (a special case of) Ichino's result, we introduce some notation. Let v denote a typical place of $\mathbb{Q}$ and p a typical prime number. Let $G = \mathrm{PGL}(2)/\mathbb{Q}$, $G_v = G(\mathbb{Q}_v)$, $K_\infty = \mathrm{SO}(2)/\{\pm 1\}$, $K_p = G(\mathbb{Z}_p)$, and $G_{\mathbb{A}} = G(\mathbb{A}) = \prod'_v G_v$, where $\mathbb{A} = \prod'_v \mathbb{Q}_v$ is the adele ring of $\mathbb{Q}$. Regard ϕ and f as pure tensors $\phi = \otimes\phi_v$ and $f = \otimes f_v$ in (factorizable) cuspidal automorphic representations $\pi_\phi = \otimes\pi_{\phi,v}$ and $\pi_f = \otimes\pi_{f,v}$ of $G_{\mathbb{A}} = \prod' G_v$. Set $\bar{f}_v = \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \cdot f_v$ and $\bar{f} = \otimes\bar{f}_v$. Then $f_p = \bar{f}_p$ for all (finite) primes p. Although the vectors ϕ_v and f_v are defined only up to a nonzero scalar multiple, the matrix coefficients

$$\Phi_{\phi,v}(g_v) = \frac{\langle g_v \cdot \phi_v, \phi_v \rangle}{\langle \phi_v, \phi_v \rangle}, \quad \Phi_{f,v}(g_v) = \frac{\langle g_v \cdot f_v, f_v \rangle}{\langle f_v, f_v \rangle}, \quad \Phi_{\bar{f},v}(g_v) = \frac{\langle g_v \cdot \bar{f}_v, \bar{f}_v \rangle}{\langle \bar{f}_v, \bar{f}_v \rangle}$$

are well-defined; here g_v belongs to G_v and $\langle , \rangle_v$ denotes the (unique up to a scalar) G_v-invariant hermitian pairings on the irreducible admissible self-contragredient representations $\pi_{\phi,v}$ and $\pi_{f,v}$. Let dg_v denote the Haar measure on the group G_v with respect to which $\mathrm{vol}(K_v) = 1$. Define the local integrals

$$I_v = \int_{G_v} \Phi_{\phi,v}(g_v)\Phi_{f,v}(g_v)\Phi_{\bar{f},v}(g_v)\, dg_v$$

and the normalized local integrals

$$\tilde{I}_v = \left(\frac{\zeta_v(2)^3}{\zeta_v(2)} \frac{L_v(\frac{1}{2}, \phi \times f \times f)}{L_v(1, \mathrm{ad}\,\phi) L_v(1, \mathrm{ad}\, f)^2} \right)^{-1} I_v, \tag{30}$$

Finally, recall from §2 that $\xi(s) = \Gamma_{\mathbb{R}}(s)\zeta(s)$, where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $\zeta(s)$ is the Riemann zeta function.

Theorem 4.3 (Ichino)**.** *We have $\tilde{I}_v = 1$ for all but finitely-many places v, and the ratio of integrals*

$$\frac{\left| \frac{\xi(2)^{-1}}{[\Gamma:\Gamma_0(q)]} \int_{\Gamma_0(q)\backslash\mathbb{H}} \phi |f|^2 y^k \, \frac{dx\, dy}{y^2} \right|^2}{\xi(2)^{-1} \int_{\Gamma\backslash\mathbb{H}} |\phi|^2 \, \frac{dx\, dy}{y^2} \left(\frac{\xi(2)^{-1}}{[\Gamma:\Gamma_0(q)]} \int_{\Gamma_0(q)\backslash\mathbb{H}} |f|^2 y^k \, \frac{dx\, dy}{y^2} \right)^2}$$

equals the ratio of L-values (times normalized local integrals)

$$\frac{\xi(2)^{-1}}{2^3} \frac{\xi(2)^3}{\xi(2)} \frac{\Lambda(\frac{1}{2}, \phi \times f \times f)}{\Lambda(1, \mathrm{ad}\,\phi)\Lambda(1, \mathrm{ad}\, f)^2} \prod_v \tilde{I}_v.$$

Proof. See [16, Theorem 1.1, Remark 1.3]. We have taken into account the relation between classical modular forms and automorphic forms on the adele group $G_{\mathbb{A}}$ (see Gelbart [9]) and the comparison (see for instance Vignéras [41, §III.2]) between the Poincaré measure on the upper half-plane and the Tamagawa measure on $G_{\mathbb{A}}$. □

We know by work of Harris and Kudla [12], Gross and Kudla [11], Watson [42], Ichino [17], and Ichino and Ikeda [17] that $\tilde{I}_\infty = 1$ and $\tilde{I}_p = 1$ for all primes p that do not divide the level q. We contribute the following computation, from which Theorem 4.1 follows upon cancelling common factors in the statement of Theorem 4.3.

Lemma 4.4. *Let p be a prime divisor of the squarefree level q. Then $\tilde{I}_p = 1/p$.*

Before embarking on the proof, let us introduce some notation and recall formulas for the matrix coefficients $\Phi_{\phi,p}$ and $\Phi_{f,p}$. Let $G_p = \mathrm{PGL}_2(\mathbb{Q}_p)$, let $K_p = \mathrm{PGL}_2(\mathbb{Z}_p)$, and let A_p be the subgroup of diagonal matrices in G_p. Recall the Cartan decomposition $G_p = K_p A_p K_p$. For $y \in \mathbb{Q}_p^*$ we write $a(y) = \left(\begin{smallmatrix} y & \\ & 1 \end{smallmatrix}\right) \in A_p$.

The representation $\pi_{\phi,p}$ is unramified principal series with Satake parameters $\alpha_\phi(p)$ and $\beta_\phi(p)$; for clarity we write simply $\alpha = \alpha_\phi(p)$ and $\beta = \beta_\phi(p)$ in this proof. The vector ϕ_p lies on the unique K_p-fixed line in $\pi_{\phi,p}$. The matrix coefficient $\Phi_{f,p}$ is bi-K_p-invariant, so by the Cartan decomposition we need only specify $\Phi_{\phi,p}(a(p^m))$ for $m \geq 0$, which is given by the Macdonald formula [3, Theorem 4.6.6]

$$\Phi_{\phi,p}(a(p^m)) = \frac{1}{1+p^{-1}} p^{-m/2} \left[\alpha^m \frac{1 - p^{-1}\frac{\beta}{\alpha}}{1 - \frac{\beta}{\alpha}} + \beta^m \frac{1 - p^{-1}\frac{\alpha}{\beta}}{1 - \frac{\alpha}{\beta}} \right]. \tag{31}$$

The representation $\pi_{f,p}$ is an unramified quadratic twist of the Steinberg representation of G_p. The vector f_p lies on the unique I_p-fixed line in $\pi_{f,p}$, where I_p is the Iwahori subgroup of K_p consisting of matrices that are upper-triangular mod p. Thus to determine $\Phi_{f,p}$, we need only specify the values it takes on representatives for the double coset space $I_p \backslash G_p / I_p$, whose structure we now recall following [10, §7] (see also [17, §7] for a similar discussion). Define the elements

$$w_1 = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \quad w_2 = \begin{pmatrix} & p^{-1} \\ p & \end{pmatrix}, \quad \omega = \begin{pmatrix} & 1 \\ p & \end{pmatrix}$$

of G_p. Note that since $G_p = \mathrm{PGL}_2(\mathbb{Q}_p)$, we have $w_1^2 = w_2^2 = \omega^2 = 1$. For w in the group $W_a = \langle w_1, w_2 \rangle$ generated by w_1 and w_2, let $\lambda(w)$ be the length of the shortest string expressing w in the alphabet $\{w_1, w_2\}$, so that $\lambda(w_1) = \lambda(w_2) = 1$. Extend λ to the group $\tilde{W} = \langle w_1, w_2, \omega \rangle$ via the formula $\lambda(\omega^i w) = \lambda(w)$ when $w \in W_a$, so that in particular $\lambda(\omega) = 0$. We have a Bruhat decomposition $G_p = \sqcup_{w \in \tilde{W}} I_p w I_p$; unwinding the definitions, this reads more concretely as

$$G_p = \left(\sqcup_{n \in \mathbb{Z}} I_p \begin{pmatrix} p^n & \\ & 1 \end{pmatrix} I_p \right) \sqcup \left(\sqcup_{n \in \mathbb{Z}} I_p w_1 \begin{pmatrix} p^n & \\ & 1 \end{pmatrix} I_p \right),$$

but we shall not adopt this perspective. With our normalization of measures we have $\mathrm{vol}(I_p w I_p) = (p+1)^{-1} p^{\lambda(w)}$. Suppose temporarily that $\pi_{f,p}$ is (the *trivial* twist of) the Steinberg representation. The matrix coefficient $\Phi_{f,p}$ is bi-I_p-invariant and given by

$$\Phi_{f,p}(\omega^j w) = (-1)^j (-p^{-1})^{\lambda(w)}$$

for all $j \in \{0, 1\}$ and $w \in W_a$. In particular

$$\Phi_{f,p}(\omega^j w)^2 = (p^{-2})^{\lambda(w)}. \tag{32}$$

In the general case that $\pi_{f,p}$ is a possibly-nontrivial unramified quadratic twist of Steinberg, the formula (32) for the *squared* matrix coefficient still holds.

Proof of Lemma 4.4. Having recalled the formulas above, we see that

$$\begin{aligned} I_p &= \int_{G_p} \Phi_{\phi,p}(g) \Phi_{f,p}(g)^2 \, dg = \sum_{w \in \tilde{W}} \mathrm{vol}(I_p w I_p) \Phi_{\phi,p}(w) (p^{-2})^{\lambda(w)} \\ &= (p+1)^{-1} \sum_{w \in \tilde{W}} \Phi_{\phi,p}(w) (p^{-1})^{\lambda(w)}, \end{aligned} \tag{33}$$

where $\Phi_{\phi,p}$ is given by (31). The evaluation of the Poincaré series

$$\sum_{w\in\tilde{W}} t^{\lambda(w)} = 2\frac{1+t}{1-t}, \tag{34}$$

where t is an indeterminate, is asserted and used in [17, §7], but we need a slightly more precise result here. For $w \in \tilde{W}$ let us write $\mu(w)$ for the unique nonnegative integer with the property that $K_p w K_p = K_p a(p^{\mu(w)}) K_p$. Then we claim that for indeterminates x, t we have the relation of formal power series

$$\sum_{w\in\tilde{W}} x^{\mu(w)} t^{\lambda(w)} = \frac{(1+x)(1+t)}{1-xt}. \tag{35}$$

Note that we recover (34) upon taking $x = 1$. To prove (35), observe that since $\omega w_1 = w_2\omega$ and $\omega^2 = 1$, every element w of $\tilde{W}$ is of the form $u_{abn} = \omega^a (w_1 w_2)^n w_1^b$ or $v_{abn} = \omega^a (w_2 w_1)^n w_2^b$ for some $a \in \{0, 1\}$, $b \in \{0, 1\}$, and $n \in \mathbb{Z}_{\geq 0}$. Computing u_{abn} and v_{abn} explicitly to be

$$u_{00n} = \begin{pmatrix} p^n & \\ & p^{-n} \end{pmatrix}, \quad u_{01n} = \begin{pmatrix} & p^n \\ p^{-n} & \end{pmatrix},$$

$$u_{10n} = \begin{pmatrix} & p^{-n} \\ p^{n+1} & \end{pmatrix}, \quad u_{11n} = \begin{pmatrix} p^{-n} & \\ & p^{n+1} \end{pmatrix},$$

$$v_{00n} = \begin{pmatrix} p^{-n} & \\ & p^n \end{pmatrix}, \quad v_{01n} = \begin{pmatrix} & p^{-n} \\ p^n & \end{pmatrix},$$

$$v_{10n} = \begin{pmatrix} & p^n \\ p^{1-n} & \end{pmatrix}, \quad v_{11n} = \begin{pmatrix} p^n & \\ & p^{1-n} \end{pmatrix},$$

we see that this parametrization of $\widetilde{W}$ is unique except that $u_{a00} = v_{a00}$ for each $a \in \{0, 1\}$; furthermore, we can read off that $\mu(u_{abn}) = 2n + a$, that $\mu(v_{abn}) = 2(n + b) - a$, and that $\lambda(u_{abn}) = \lambda(v_{abn}) = 2n + b$. Thus

$$\begin{aligned}
\sum_{w\in\tilde{W}} x^{\mu(w)} t^{\lambda(w)} &= (1+x) + \sum_{\substack{b=0,1 \\ 2n+b>0}} \sum_{n\geq 0} t^{2n+b} \sum_{a=0,1} \left(x^{2n+a} + x^{2(n+b)-a}\right) \\
&= (1+x) + \sum_{\substack{b=0,1 \\ 2n+b>0}} \sum_{n\geq 0} t^{2n+b} x^{2n+b-1} \sum_{a=0,1} \left(x^{1+a-b} + x^{1+b-a}\right) \\
&= (1+x) + (1+x)^2 \sum_{m>0} t^m x^{m-1},
\end{aligned}$$

from which (35) follows upon summing the geometric series. We now combine (31), (33) and (35), noting that the series converge because $|\alpha| < p^{1/2}$ and $|\beta| < p^{1/2}$; the contributions to the formula (33) for I_p of the two terms in the formula (31) for $\Phi_{\phi,p}(a(p^m))$ are respectively

$$(p+1)^{-1}(1+p^{-1})^{-1} \frac{1 - p^{-1}\frac{\beta}{\alpha}}{1 - \frac{\beta}{\alpha}} \frac{(1 + p^{-1/2}\alpha)(1 + p^{-1})}{1 - p^{-3/2}\alpha}$$

and

$$(p+1)^{-1}(1+p^{-1})^{-1} \frac{1 - p^{-1}\frac{\alpha}{\beta}}{1 - \frac{\alpha}{\beta}} \frac{(1 + p^{-1/2}\beta)(1 + p^{-1})}{1 - p^{-3/2}\beta}.$$

Summing these fractions by cross-multiplication and then simplifying, we obtain

$$I_p = p^{-1}(1 - p^{-1}) \frac{(1 + \alpha p^{-1/2})(1 + \beta p^{-1/2})}{(1 - \alpha p^{-3/2})(1 - \beta p^{-3/2})}.$$

Recall the definition (30) of $\tilde{I}_p$. The local L-factors are given by (see [42, §3.1])

$$L_p(1, \operatorname{ad} f) = \zeta_p(2), \quad L_p(1, \operatorname{ad} \phi) = [(1-\alpha^2 p^{-1})(1-p^{-1})(1-\beta^2 p^{-1})]^{-1},$$

$$L_p(\tfrac{1}{2}, \phi \times f \times f) = [(1-\alpha p^{-1/2})(1-\beta p^{-1/2})(1-\alpha p^{-3/2})(1-\beta p^{-3/2})]^{-1},$$

thus the normalized local integral $\tilde{I}_p$ is

$$\tilde{I}_p = p^{-1}(1-p^{-1}) \frac{(1-\alpha p^{-1/2})(1-\beta p^{-1/2})(1+\alpha p^{-1/2})(1+\beta p^{-1/2})}{(1-\alpha^2 p^{-1})(1-p^{-1})(1-\beta^2 p^{-1})} = p^{-1},$$

as asserted. ☐

5. Proof of Theorem 1.3

We combine Theorem 3.1 and Theorem 4.1 with Soundararajan's weak subconvex bounds [38] to complete the proof of Theorem 1.3. Fix a positive even integer k. Let f be a newform of weight k and squarefree level q. Fix a Maass eigencuspform or incomplete Eisenstein series ϕ. We will show that the "discrepancy"

$$D_f(\phi) := \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)}.$$

tends to 0 as $q \to \infty$, thereby fulfilling the criterion of Lemma 1.10, by combining the complementary estimates for $D_f(\phi)$ provided below by Proposition 5.2 and Proposition 5.3.

Lemma 5.1. *The quantities $M_f(x)$ and $R_f(x)$ (15) appearing in the statement of Theorem 3.1 satisfy the estimates*

$$M_f(q) \ll_{k,\varepsilon} \log(eq)^{1/6+\varepsilon} L(\operatorname{ad} f, 1)^{1/2}, \quad R_f(q) \ll_{k,\varepsilon} \frac{\log(eq)^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)} \ll_k \log(eq)^{\varepsilon}$$

Proof. The bound for $M_f(q)$ follows from the proof of [14, Lemma 3] with "k" replaced by "q," noting that $\lambda_f(p)^2 \leq 1 + \lambda_f(p^2)$ for *all* primes p. The bound for $R_f(q)$ follows from the arguments of [38, Example 1], [14, Lemma 1] with "k" replaced by "q" and the lower bound (6) for $L(\operatorname{ad} f, 1)$. ☐

Proposition 5.2. *We have $D_f(\phi) \ll_{k,\phi,\varepsilon} \log(eq)^{1/12+\varepsilon} L(\operatorname{ad} f, 1)^{1/4}$.*

Proof. Follows immediately from Theorem 3.1 and Lemma 5.1. ☐

Proposition 5.3. *We have $D_f(\phi) \ll_{k,\phi,\varepsilon} \log(eq)^{-\delta+\varepsilon} L(\operatorname{ad} f, 1)^{-1}$, where $\delta = 1/2$ if ϕ is a Maass eigencuspform and $\delta = 1$ if ϕ is an incomplete Eisenstein series.*

Proof. If ϕ is a Maass eigencuspform, then the analytic conductor of $\phi \times f \times f$ is $\asymp k^4 q^4 \asymp_k q^4$, so Theorem 4.1 and the arguments of Soundararajan [38, Example 2] with "k" replaced by "q" show that

$$\left| \frac{\mu_f(\phi)}{\mu_f(1)} \right|^2 \ll_{k,\phi} \frac{L(\phi \times f \times f, \frac{1}{2})}{q \cdot L(\operatorname{ad} f, 1)^2} \ll_{\varepsilon} \frac{1}{\log(eq)^{1-\varepsilon} L(\operatorname{ad} f, 1)^2}.$$

If $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series, then the unfolding method as in Lemma 3.5 and the bound for $R_f(q)$ given by Lemma 5.1 show that

$$\begin{aligned} \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} &= \frac{2\pi^2}{q} \int_{(1/2)} \Psi^\wedge(s) \left(\frac{q}{4\pi}\right)^s \frac{\Gamma(s+k-1)}{\Gamma(k)} \frac{\zeta(s)}{\zeta(2s)} \frac{L(\operatorname{ad} f, s)}{L(\operatorname{ad} f, 1)} \frac{ds}{2\pi i} \\ &\ll_{k,\phi} R_f(q) \ll_{k,\varepsilon} \frac{\log(eq)^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}. \end{aligned}$$

☐

Proof of Theorem 1.3. By Propositions 5.2 and 5.3, there exists $\delta \in \{1/2, 1\}$ such that

$$D_f(\phi) \ll_{k,\phi,\varepsilon} \min\left(\log(eq)^{-\delta+\varepsilon} L(\mathrm{ad}\, f, 1)^{-1}, \log(eq)^{1/12+\varepsilon} L(\mathrm{ad}\, f, 1)^{1/4}\right);$$

it follows by the argument of [14, §3] with "k" replaced by "q" that $D_f(\phi) \to 0$ as $q \to \infty$. □

References

[1] A. O. L. Atkin and J. Lehner. Hecke operators on $\Gamma_0(m)$. *Math. Ann.*, 185:134–160, 1970.

[2] S. Böcherer and R. Schulze-Pillot. On the central critical value of the triple product L-function. In *Number theory (Paris, 1993–1994)*, volume 235 of *London Math. Soc. Lecture Note Ser.*, pages 1–46. Cambridge Univ. Press, Cambridge, 1996.

[3] Daniel Bump. *Automorphic forms and representations*, volume 55 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[4] Pierre Deligne. Formes modulaires et reprsentations ℓ-adiques. *Séminaire Bourbaki, Vol. 1968/1969, Exp. 347-363*, 179:139–172, 1971.

[5] Pierre Deligne. La conjecture de Weil. I. *Inst. Hautes Études Sci. Publ. Math.*, 43:273–307, 1974.

[6] John Friedlander and Henryk Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.

[7] Paul B. Garrett. Decomposition of Eisenstein series: Rankin triple products. *Ann. of Math. (2)*, 125(2):209–235, 1987.

[8] Stephen Gelbart and Hervé Jacquet. A relation between automorphic representations of GL(2) and GL(3). *Ann. Sci. École Norm. Sup. (4)*, 11(4):471–542, 1978.

[9] Stephen S. Gelbart. *Automorphic forms on adèle groups*. Princeton University Press, Princeton, N.J., 1975. Annals of Mathematics Studies, No. 83.

[10] Roger Godement and Hervé Jacquet. *Zeta functions of simple algebras*. Lecture Notes in Mathematics, Vol. 260. Springer-Verlag, Berlin, 1972.

[11] Benedict H. Gross and Stephen S. Kudla. Heights and the central critical values of triple product L-functions. *Compositio Math.*, 81(2):143–209, 1992.

[12] Michael Harris and Stephen S. Kudla. The central critical value of a triple product L-function. *Ann. of Math. (2)*, 133(3):605–672, 1991.

[13] Jeffrey Hoffstein and Paul Lockhart. Coefficients of Maass forms and the Siegel zero. *Ann. of Math. (2)*, 140(1):161–181, 1994. With an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman.

[14] R. Holowinsky and K. Soundararajan. Mass equidistribution for hecke eigenforms. `arXiv.org:0809.1636`, 2008.

[15] Roman Holowinsky. Sieving for mass equidistribution. `arXiv.org:0809.1640`, 2008.

[16] Atsushi Ichino. Trilinear forms and the central values of triple product L-functions. *Duke Math. J.*, 145(2):281–307, 2008.

[17] Atsushi Ichino and Tamutsu Ikeda. On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture. *Geom. Funct. Anal.*, 19(5):1378–1425, 2010.

[18] Henryk Iwaniec. *Topics in classical automorphic forms*, volume 17 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.

[19] Henryk Iwaniec. *Spectral methods of automorphic forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2002.

[20] E. Kowalski, P. Michel, and J. VanderKam. Rankin-Selberg L-functions in the level aspect. *Duke Math. J.*, 114(1):123–191, 2002.

[21] Shin-ya Koyama. Equidistribution of Eisenstein series in the level aspect. *Comm. Math. Phys.*, 289(3):1131–1150, 2009.

[22] Elon Lindenstrauss. Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)*, 163(1):165–219, 2006.

[23] Wen Zhi Luo and Peter Sarnak. Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$. *Inst. Hautes Études Sci. Publ. Math.*, (81):207–237, 1995.

[24] Wenzhi Luo and Peter Sarnak. Mass equidistribution for Hecke eigenforms. *Comm. Pure Appl. Math.*, 56(7):874–891, 2003. Dedicated to the memory of Jürgen K. Moser.

[25] Mohan Nair. Multiplicative functions of polynomial values in short intervals. *Acta Arith.*, 62(3):257–269, 1992.

[26] Paul Nelson. Mass equidistribution of Hilbert modular eigenforms. `http://www.its.caltech.edu/~pnelson/hque-hilbert.pdf`, 2010.

[27] I. Piatetski-Shapiro and Stephen Rallis. Rankin triple L functions. *Compositio Math.*, 64(1):31–115, 1987.

[28] R. A. Rankin. Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. II. *Proc. Cambridge Philos. Soc.*, 35:351–372, 1939.

[29] Zeév Rudnick. On the asymptotic distribution of zeros of modular forms. *Int. Math. Res. Not.*, (34):2059–2074, 2005.

[30] Zeév Rudnick and Peter Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Comm. Math. Phys.*, 161(1):195–213, 1994.

[31] Peter Sarnak. Arithmetic quantum chaos. In *The Schur lectures (1992) (Tel Aviv)*, volume 8 of *Israel Math. Conf. Proc.*, pages 183–236. Bar-Ilan Univ., Ramat Gan, 1995.

[32] Peter Sarnak. Recent Progress on QUE. `http://www.math.princeton.edu/sarnak/SarnakQUE.pdf`, 2009.

[33] Atle Selberg. Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist. *Arch. Math. Naturvid.*, 43:47–50, 1940.

[34] Atle Selberg. On the estimation of Fourier coefficients of modular forms. In *Proc. Sympos. Pure Math., Vol. VIII*, pages 1–15. Amer. Math. Soc., Providence, R.I., 1965.

[35] J.-P. Serre. *A course in arithmetic*. Springer-Verlag, New York, 1973. Translated from the French, Graduate Texts in Mathematics, No. 7.

[36] Goro Shimura. *Introduction to the arithmetic theory of automorphic functions*. Publications of the Mathematical Society of Japan, No. 11. Iwanami Shoten, Publishers, Tokyo, 1971. Kanô Memorial Lectures, No. 1.

[37] Goro Shimura. On the holomorphy of certain Dirichlet series. *Proc. London Math. Soc. (3)*, 31(1):79–98, 1975.

[38] K. Soundararajan. Weak subconvexity for central values of L-functions. `arXiv.org:0809.1635`, 2008.

[39] K. Soundararajan. Quantum unique ergodicity for SL_2(Z)\H. *ArXiv e-prints*, January 2009.

[40] K. Soundararajan. Arizona winter school project outline for quantum unique ergodicity and number theory. `http://math.arizona.edu/~swc/aws/10/2010SoundararajanOutlineProjects.pdf`, 2010.

[41] Marie-France Vignéras. *Arithmétique des algèbres de quaternions*, volume 800 of *Lecture Notes in Mathematics*. Springer, Berlin, 1980.

[42] Thomas C. Watson. Rankin triple products and quantum chaos. `arXiv.org:0810.0425`, 2008.