Some New Aspects of Mass Equidistribution

Thesis by

Paul Nelson

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2011
(Defended May 26, 2011)

to my parents

Acknowledgments

I would like to extend my deepest thanks to my advisor, Professor Dinakar Ramakrishnan, for his generous and enthusiastic support and guidance and for the wealth of mathematics he has guided me toward during my time at Caltech, both as a student in his many topics courses and as research mentee. I would like to thank the Mathematics Department at Caltech for the excellent working environment, support, and general helpfulness. I would like to thank my friends and family for their friendship and encouragement.

Abstract

This thesis presents two new results concerning the limiting behavior of families of automorphic forms. The work is presented in a sequence of chapters. The first, "Mass equidistribution of Hilbert modular eigenforms," has been accepted for publication in the *Ramanujan Journal* (Springer), while the second, "Equidistribution of cusp forms in the level aspect,"[1] has been accepted for publication in the *Duke Mathematical Journal* (Duke University Press). Some minor differences exist between these chapters and the papers they represent. The abstracts of the accepted versions of these papers follow.

1. Let $\mathbb{F}$ be a totally real number field, and let f traverse a sequence of nondihedral holomorphic eigencuspforms on $\mathrm{GL}_2/\mathbb{F}$ of weight $(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$, trivial central character, and full level. We show that the mass of f equidistributes on the Hilbert modular variety as $\max(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty$.

   Our result answers affirmatively a natural analogue of a conjecture of Rudnick and Sarnak (1994). Our proof generalizes the argument of Holowinsky-Soundararajan (2008) who established the case $\mathbb{F} = \mathbb{Q}$. The essential difficulty in doing so is to adapt Holowinsky's bounds for the Weyl periods of the equidistribution problem in terms of manageable shifted convolution sums of Fourier coefficients to the case of a number field with nontrivial unit group.

2. 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 in 2002 by Kowalski, Michel, and VanderKam [36] in the spirit of a conjecture of Rudnick and Sarnak [52] made in 1994.

Our proof follows the strategy of Holowinsky and Soundararajan [25] who showed in 2008 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, a refinement of his bounds for shifted sums, 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

Chapter 1

Introduction

The basic problem addressed in this work is the study of the limiting behavior of families of automorphic forms and special values of L-functions. Automorphic forms and their L-functions, which generalize the classical zeta function of Riemann, are fundamental in modern number theory.

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 $Y_0(q) = \Gamma_0(q)\backslash\mathbb{H}$. Our starting point is the recent proof by Holowinsky and Soundararajan [25] that newforms of large weight k and fixed level $q = 1$ have equidistributed mass with respect to the hyperbolic area measure, answering affirmatively a natural variant[1] of the *quantum unique ergodicity* conjecture of Rudnick and Sarnak [52].

Theorem 1.0.1 (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 $Y_0(q)$.*

[1]as spelled out by Luo and Sarnak [42]; we refer to Sarnak [53, 54] and the references in [25] 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 $C_c(X)$ and writing 1 for the constant function.

We prove two new variants of the Holowinsky-Soundararajan result by suitably adapting their method and tackling some new subtleties that arise. Before stating our main results, let us highlight two perspectives from which the study of limiting behavior of the masses of automorphic forms is natural and interesting. First, it is analogous to a fundamental problem in quantum chaos, which concerns more generally the limiting behavior as $\lambda \to \infty$ of eigenfunctions ϕ

$$(\Delta + \lambda)\phi = 0 \tag{1.1}$$

of the Laplacian Δ on a compact Riemannian manifold M for which the geodesic flow is chaotic (see [53]). Here the geodesic flow on M is regarded as the Hamiltonian flow of a chaotic classical mechanical system, the Laplacian $\Delta \circlearrowright L^2(M)$ as the Hamiltonian operator for the corresponding quantized system, and the eigenfunction ϕ (normalized so that $\int |\phi|^2 = 1$) as the wave function for a quantum particle on M of energy λ whose position is described in the Copenhagen interpretation of quantum mechanics by the probability density $|\phi|^2$. In suitable units the Schrödinger equation for stationary states reads $(\hbar^2\Delta + \lambda)\phi = 0$, so studying ϕ in (1.1) as $\lambda \to \infty$ is akin to considering the semiclassical limit $\hbar \to 0$ of the quantization of the geodesic flow.

Among several questions that one can ask we single out that of the behavior of the densities $|\phi|^2$ for particles of high energy $\lambda \to \infty$. A fundamental result in this direction is the *quantum ergodicity* theorem of Schnirelman, Colin de Verdière, and Zelditch [56, 6, 73], which asserts that if the geodesic flow on the unit cotangent bundle of M is ergodic, then for any sequence (ϕ_n) with $\lambda_n \to \infty$ there exists a full-density subsequence (ϕ_{n_k}) such that the $|\phi_{n_k}|^2$ equidistribute.[3] In the particular case that M is negatively curved, the *quantum unique ergodicity* (QUE) conjecture of Rudnick and Sarnak [52] predicts that the *full* sequence of $|\phi_n|^2$ equidistributes with respect to the volume measure on M as $\lambda \to \infty$.

The QUE conjecture is considered difficult and there has been little progress for general M, but for certain special M that arise from arithmetic considerations (such as the modular curve or the Hilbert modular varieties) there has been significant progress on QUE and related questions [54, 41, 40, 65, 63, 25]. Such arithmetic manifolds arise as quotients of symmetric spaces by arithmetic groups and are characterized by the presence of additional symmetry in the form of a large commuting family $\mathbb{T}$ of correspondences that commute with the algebra $\mathcal{D}$ of invariant differential operators, thereby providing a powerful tool for the study of common eigenfunctions of $\mathbb{T}$ and $\mathcal{D}$. One may hope that such arithmetic instances of QUE provide tractable and yet

[3] in a more precise sense than we describe here; see the introduction to [52]

representative model cases for the more general problem (see [53]).

A second motivation for our considerations arises from their connection to central problems in the analytic theory of L-functions. Watson [70] showed that for $M = \mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$ (as well as other "arithmetic surfaces" $\Gamma\backslash\mathbb{H}$), the Weyl periods for the equidistribution problem posed by QUE are essentially products of central values $L(\frac{1}{2})$ of automorphic L-functions $L(s)$ of degree at most 6; a similar relation holds over totally real fields (see §2.3.2) and for newforms of varying level (see remark 15). The generalized Riemann hypothesis (GRH) for such $L(s)$, which asserts that the nontrivial zeros of $L(s)$ lie on the line $\mathrm{Re}(s) = \frac{1}{2}$, would imply sufficiently strong bounds on $L(\frac{1}{2})$ to establish the QUE conjecture for M. But the bounds on $L(\frac{1}{2})$ demanded by QUE are considerably more tractable than those implied by the GRH (let alone the GRH itself), and so provide accessible problems on which to develop new techniques.

We now give somewhat informal statements of our main results. In chapter 2, we generalize theorem 1.0.1 to an arbitrary totally real number field $\mathbb{F}$, where the main technical challenge for $[\mathbb{F} : \mathbb{Q}] > 1$ is presented by the infinite unit group. This result specializes to theorem 1.0.1 in the case $\mathbb{F} = \mathbb{Q}$.

Theorem 1.0.2 (Mass equidistribution for Hilbert modular eigenforms in the max-weight aspect)**.** *Let $\mathbb{F}$ be a totally real number field, and let f traverse a sequence of full-level nondihedral holomorphic eigencuspforms on* $\mathrm{PGL}(2)/\mathbb{F}$ *with any weight component of f tending to ∞. Then the mass of f equidistributess with respect to the invariant measure on the appropriate adelic quotient of* $\mathrm{PGL}(2)/\mathbb{F}$*.*

Kowalski, Michel, and VanderKam [36, Conj 1.5] formulated an analogue of theorem 1.0.1 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 1.0.3 (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 under π_q of the mass of f equidistributes with respect to the Poincaré measure on* $Y_0(1)$*.*

In chapter , we prove the squarefree level case of Conjecture 1.0.3.

Theorem 1.0.4 (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 the pushforward under π_q of the mass of f to* $Y_0(1)$ *equidistributes with respect to the Poincaré measure on* $Y_0(1)$*.*

The main technical difficulties here are to find a suitable generalization of Holowinsky's unfolding method for forms of increasing level, to improve his bounds for shifted convolution sums in their dependence on the size of the shift with respect to the size of the summation interval, and to generalize Watson's formula relating the integral of the product of three newforms of the same squarefree to the central L-value of their triple product L-function to the case of triples of newforms of possibly varying squarefree level.

Having stated informally our main results, we now survey the ideas involved in their proofs. In chapter 2, we consider nondihedral holomorphic Hilbert modular eigencuspforms f on $\mathrm{PGL}_2/\mathbb{F}$ of weight $(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$ and full level, the equidistribution of whose mass we seek on the (in general, non-connected) Hilbert modular variety Y. The basic strategy, as in many equidistribution problems, is to study the "Weyl periods" $\int \phi |f|^2$ as ϕ traverses a *convenient spanning set* of functions on Y, analogous to how one uses the exponentials $\mathbb{R}/\mathbb{Z} \ni x \mapsto e^{2\pi i n x}$ to prove the equidistribution of the fractional parts of αk ($k \in \mathbb{N}$) for $\alpha \in \mathbb{R} - \mathbb{Q}$.

Indeed, theorem 2.1.1 follows as soon as one can establish (2.1) for each element ϕ of a set the uniform closure of whose span contains $C_c(Y)$. Such a spanning set is furnished by the Maass eigencuspforms and the incomplete Eisenstein series, as defined in §2.2.8. To highlight the essential difficulties let us suppose in this section that ϕ is a Maass eigencuspform. Then $\int \phi = 0$, so to establish (2.1) we must show that

$$\frac{\int \phi |f|^2}{\int |f|^2} \to 0 \quad \text{ as } \max(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty, \tag{1.2}$$

where the rate of convergence is allowed to depend upon ϕ.

Take $\mathbb{F} = \mathbb{Q}$ and f of weight k for now. Holowinsky and Soundararajan established (1.2) by a remarkable synthesis of their independent efforts [24, 66], which we now recall briefly, saving a more detailed discussion for §2.3 and referring to the lucid expositions of [25, 54, 64] for further motivation and details. Watson's formula [70] and work of Gelbart-Jacquet [14] and Hoffstein-Lockhart-Goldfeld-Lieman [22] imply (see [25, Lem 2]) that

$$\frac{\int \phi |f|^2}{\int |f|^2} \approx_\phi \frac{|L(\phi \times \mathrm{ad}\, f, \frac{1}{2})|^{1/2}}{k^{1/2}} \exp\left(- \sum_{p \le k} \frac{1}{p} \lambda(p^2) \right), \tag{1.3}$$

where $L(\cdot)$ denotes the finite part of the L-function indicated above, $\approx_\phi$ denotes equality up to multiplication by a bounded power of $\log\log(k)$ times a constant depending upon ϕ, and $\lambda(n)$ is the nth Fourier coefficient of f normalized so that the Deligne bound reads $|\lambda(p)| \le 2$. Soundararajan proves a "weak subconvexity" bound for the central values of quite general L-functions satisfying a "weak Ramanujan hypothesis," specializing in the present circumstances to $|L(\phi \times \mathrm{ad}\, f, \frac{1}{2})| \ll k/\log(k)^{1-\varepsilon}$ for any $\varepsilon > 0$, which implies (1.2) provided that

$$\frac{\sum_{p \le k} \frac{1}{p}\lambda(p^2)}{\sum_{p\le k} \frac{1}{p}} \ge -1/2 + \delta + o_{k\to\infty}(1) \quad \text{for some fixed } \delta > 0. \tag{1.4}$$

By considering Fourier expansions at the cusps of the modular curve and bounding the sums (described below in more detail) that arise, Holowinsky proves (following the reformulation of Iwaniec [30])

$$\frac{\int \phi |f|^2}{|f|^2} \ll_{\phi,\varepsilon} \log(k)^{\varepsilon} \exp\left(- \sum_{p \le k} \frac{1}{p} (|\lambda(p)| - 1)^2 \right), \tag{1.5}$$

which implies (1.2) provided that

$$\frac{\sum_{p \le k} \frac{1}{p}(|\lambda(p)| - 1)^2}{\sum_{p\le k} \frac{1}{p}} \ge \delta + o_{k\to\infty}(1) \quad \text{for some fixed } \delta > 0. \tag{1.6}$$

In summary, Soundararajan succeeds unless typically $\lambda(p^2) \lessapprox -1/2$, while Holowinsky succeeds unless typically $|\lambda(p)| \approx 1$ (in the harmonically weighted sense taken over $p \le k$); the identity $\lambda(p)^2 = \lambda(p^2) + 1$ shows that

$$\lambda(p^2) \lessapprox -1/2 \implies |\lambda(p)| \lessapprox \sqrt{1/2} \quad \text{and} \quad |\lambda(p)| \approx 1 \implies \lambda(p^2) \approx 0,$$

so in all cases at least one of their approaches succeeds.

The basic ideas underlying our proof when $\mathbb{F}$ is totally real are the same as those just described in the case $\mathbb{F} = \mathbb{Q}$; the generalization is a nontrivial and yet purely technical matter, requiring no fundamental reworking of the overall strategy. As we shall explain in §2.3, the only part of the $\mathbb{F} = \mathbb{Q}$ argument that does not generalize transparently is Holowinsky's proof of (1.5). His argument amounts to

1. bounding $\int \phi |f|^2 / \int |f|^2$ from above in terms of the "shifted sums"

$$X^{-1} \sum_{n \in \mathbb{Z} \cap [1,X]}^{\text{smooth}} \lambda(n)\lambda(n+l), \tag{1.7}$$

   where $l \neq 0$ is a small integer and $X \approx k$, and

2. bounding the shifted sums (1.7); a reformulation [30] of the bound that Holowinsky obtains is

$$X^{-1} \sum_{n, n+l \in \mathbb{Z}\cap[1,X]} |\lambda(n)\lambda(n+l)| \ll \tau(l) \log(k)^{\varepsilon} \prod_{p \le k} \left(1 + \frac{2(|\lambda(p)| - 1)}{p} \right), \tag{1.8}$$

   which is roughly the square of the bound one would expect for $X^{-1} \sum |\lambda(n)|$ and so may be understood as asserting the independence of the random variables $n \mapsto |\lambda(n)|$, $n \mapsto |\lambda(n+l)|$ owing to the independence of the prime factorizations of n and $n+l$ and the multiplicativity of λ. The novelty in his argument is that he does not exploit cancellation in the sums (1.7) that one would expect to arise from the independent variation in sign of $\lambda(n)$ and $\lambda(n+l)$ for varying n and fixed $l \neq 0$; his motivation for doing so came from the expectation that the $\lambda(p)$ follow the Sato-Tate distribution, which suggests that $X^{-1}\sum|\lambda(n)| \ll \log(X)^{-\delta}$ for some small $\delta > 0$. See [42, 25, 54, 64] and especially [23] for further discussion.

Now let $[\mathbb{F} : \mathbb{Q}] = d$ and take f of weight $(k_1, \ldots, k_d)$. The most naïve higher-dimensional generalization of Holowinsky's method that we found requires one to replace X and $\mathbb{Z} \cap [1, X]$ in (1.7) by $X \approx k_1 \cdots k_d$ and $\mathfrak{o} \cap \mathcal{R}$, where $\mathfrak{o}$ is the ring of integers in $\mathbb{F}$ and $\mathcal{R}$ is the region in the totally positive quadrant of $\mathbb{F} \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R}^d$ bounded by the hyperbola $\{x_1 \cdots x_d = X\}$ and the hyperplanes $\{x_i = c\}$ for some small constant $c > 0$. Unfortunately, the volume of $\mathcal{R}$ is roughly $X \log(X)^{d-1}$, so even the most optimistic bounds along the lines of (1.8) fail to produce an estimate of the quality (1.5) because of the unaffordable factor $\log(X)^{d-1}$ when $d > 1$.

To circumvent this difficulty, we refine Holowinsky's upper bound for $\int \phi|f|^2$ by a method that when $\mathbb{F} = \mathbb{Q}$ leads (see remark 1) to the precise asymptotic expansion

$$\frac{\int \phi|f|^2}{\int |f|^2} \sim \frac{(Yk)^{-1}}{L(\mathrm{ad}\, f, 1)} \sum_{\substack{m=n+l \\ \max(m,n) \asymp Yk}} \sum \frac{\lambda_\phi(l)}{\sqrt{|l|}} \lambda_f(m) \lambda_f(n) \kappa_{\phi,\infty}\left(\frac{k-1}{4\pi} \left|\log \frac{m}{n}\right|\right), \tag{1.9}$$

where $Y \geq 1$ tends slowly to infinity with k, λ_ϕ, and λ_f are the normalized Fourier coefficients of ϕ and f respectively, $\kappa_{\phi,\infty}(y) = 2y^{1/2}K_{ir}(2\pi y)$ for $y > 0$ if $\frac{1}{4} + r^2$ is the Laplace eigenvalue of ϕ, and the sum is taken over triples $(l, m, n) \in \mathbb{Z}^3$ for which $0 \neq |l| < Y^{1+\varepsilon}$, $m > 0, n > 0$, $m - n = l$ and $\max(m, n) \asymp Yk$ (with the last condition imposed by a normalized smooth truncation).

We exploit (in Lemma 2.4.3 and Corollary 2.4.4; see also remark 2) what amounts to the overwhelming decay of the Bessel factor $\kappa_{\phi,\infty}(\cdots)$ in the higher-dimensional generalization of (1.9) when m, n lie in the outskirts of the region $\mathcal{R}$; the simple proof that we give amounts to some amusing inequalities satisfied by the hypergeometric function and ratios of pairs of Gamma functions (see §2.8). In this way we reduce to bounding shifted sums of the form (1.7) taken over $\mathfrak{o} \cap \mathcal{R}'$ with $\mathcal{R}'$ the much smaller region bounded by the hyperbola $\{x_1 \cdots x_d = X\}$ and the hyperplanes $\{x_i = k_i Y^{1/d}/U\}$ with $X = k_1 \cdots k_d Y$ and $U = \exp(\log(X)^\varepsilon)$. The volume of $\mathcal{R}'$ is merely $\approx X \log(U)^{d-1} = X \log(X)^{\varepsilon'}$ with $\varepsilon' = (d-1)\varepsilon$, and this arbitrarily small logarithmic power $\log(X)^{\varepsilon'}$ is negligible in seeking estimates of type (1.8) and (1.5) which already contain such a factor. The rest of our argument proceeds essentially as it did for Holowinsky upon replacing his Mellin transforms on $\mathbb{R}_+^*$ by Mellin transforms on certain quotients of the idele class group of $\mathbb{F}$, although some new features do arise (e.g., when $\mathbb{F}$ has general class number we must consider Hilbert modular varieties having multiple connected components and exclude certain dihedral forms from our analysis). We elaborate on these last few paragraphs in successively greater detail in §2.3 and §2.4.

In chapter , where we consider the limiting behavior of mass of classical newforms of large level, the synthetic part of the Holowinsky-Soundararajan argument works just as well as in the weight aspect, so we highlight here four of the more substantial difficulties encountered in adapting the independent arguments of Holowinsky and Soundararajan to the level aspect.

First, it is not a priori clear how best to extend Holowinsky's unfolding trick in the presence of multiple (possibly unboundedly many) cusps, nor what should take the place of his asymptotic analysis of archimedean integrals in studying the fixed weight, large level limit; several fundamentally different approaches are possible, one of which we shall present in §3.3.1. When q is squarefree, the problem then becomes to bound sums roughly of the form[4]

$$\sum_{d|q} \sum_{n \ll dk} \lambda_f(n) \lambda_f(n + dl), \tag{1.10}$$

where again $l \neq 0$ is essentially bounded. As we now explain, the sums (1.10) differ from the sums

$$\sum_{n \ll k} \lambda_f(n) \lambda_f(n + l), \tag{1.11}$$

studied by Holowinsky in two important ways.

For one, the shifts dl are now nearly as large as the length of the interval $\approx dk$ over which we are summing.[5] Much of the existing work on bounds for such sums (see remark 13) applies only when the shift is substantially smaller than the summation interval. Holowinsky's treatment of (1.11) *does* allow shifts as large as the summation interval, but gives a bound for $\sum_{n \ll qk} \lambda_f(n) \lambda_f(n + ql)$ that involves an extraneous factor of $\tau(ql)$, which is prohibitively large (e.g., $\gg \log(q)^A$ for any A) if q has many small prime factors. In theorem 3.3.10, we refine Holowinsky's method to allows shifts as large as the summation interval with *full uniformity* in the size of the shift, e.g., without the factor $\tau(ql)$. This refinement may be of independent

[4]Here one should think of a divisor d of q as indexing the unique cusp of $\Gamma_0(q)$ of width d, where as usual the *width* of a cusp is its ramification index over the cusp ∞ for $\Gamma_0(1)$.

[5]This difficulty corresponds the fact that cusps for $\Gamma_0(q)$ may have *large* width.

interest.

Now let $\omega(q)$ denote the number of prime divisors of the squarefree integer q. Then the number of shifted sums in (1.10) is $2^{\omega(q)}$, which can be quite large.[6] In the crucial case[7] that $|\lambda_f(p)|$ is typically small for primes $p \ll qk$, our refinement of Holowinsky's method saves nearly two logarithmic powers of dk over the trivial bound $\ll dk$ for the shifted sum in (1.10) of length $\approx dk$. Thus we save very little over the trivial bound if d is a small divisor of q, and it is not immediately clear whether such savings are sufficient to produce a sufficient saving in the sum over all d. One needs here an inequality of the shape

$$\sum_{d|q} \frac{dk}{\log(dk)^{2-\varepsilon}} \ll \frac{qk}{\log(qk)^{2-\varepsilon}} \log\log(e^e q), \tag{1.12}$$

which one can interpret as saying that the divisors of any squarefree integer are well distributed in a certain sense. Indeed, if hypothetically q were to have "too many" large divisors, then the LHS of (1.12) might be large enough to swamp the small logarithmic savings, while if q were to have "too many" small divisors, then the savings for each term on the LHS might be too small to produce an overall savings. A convexity argument and a (weak form of the) prime number theorem are sufficient to establish (1.12); see Lemma 3.3.13.

Finally, the identity relating $\mu_f(\phi)$ to $L(\phi \times f \times f, \frac{1}{2})$ that Soundararajan's method takes as input is given by Watson [70] when f and ϕ are newforms of the *same* (squarefree) level. In the level aspect, the relevant Weyl periods are those for which f has large level and ϕ has fixed level, so Watson's formula does not apply. We extend Watson's result in theorem 3.4.1 by computing (Lemma 3.4.3) a p-adic integral arising in Ichino's general formula [26], specifically

$$\int_{g \in \mathrm{PGL}_2(\mathbb{Q}_p)} \frac{\langle g \cdot \phi_p, \phi_p \rangle}{\langle \phi_p, \phi_p \rangle} \frac{\langle g \cdot f_p, f_p \rangle}{\langle f_p, f_p \rangle} \frac{\langle g \cdot f_p, f_p \rangle}{\langle f_p, f_p \rangle} \, dg, \tag{1.13}$$

where ϕ_p (resp. f_p) is the newvector at p for the adelization of ϕ (resp. f) and $\langle , \rangle$ denotes a $\mathrm{PGL}(2, \mathbb{Q}_p)$-invariant Hermitian pairing on the appropriate representation space. 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. As we discuss in remark 15, our evaluation of (1.13) leads to a precise formula relating $\int \psi_1 \psi_2 \psi_3$ to $L(\frac{1}{2}, \psi_1 \times \psi_2 \times \psi_3)$ for *any* three newforms of squarefree level (and trivial central character); such an identity should be of

[6] This difficulty corresponds to the fact that $\Gamma_0(q)$ may have *many* cusps.

[7] Soundarajan's argument succeeds unless this is so.

general use in future work that exploits the connection between periods and L-values.

Chapter 2

Mass Equidistribution of Hilbert Modular Eigenforms

2.1 Introduction

2.1.1 Statement of Main Result

Let $\mathbb{F}$ be a totally real number field and f a holomorphic Hilbert modular eigencuspform on $\mathrm{PGL}_2/\mathbb{F}$ of weight $k = (k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$ and full level. The mass $|f|^2$ descends to a finite measure on the Hilbert modular variety; our aim in this chapter is to prove that the measures so obtained equidistribute with respect to the uniform measure as the weight k of f tends to ∞. Motivation for this problem, as discussed in §1, comes from its connection to quantum chaos by analogy with the *quantum unique ergodicity* conjecture of Rudnick and Sarnak [52] as well as from its connection to central problems in the analytic theory of L-functions, specifically those such as the subconvexity problem that concern the rate of growth of central L-values. Our result and its method of proof directly generalize recent work of Holowinsky and Soundararajan [25] in the case $\mathbb{F} = \mathbb{Q}$, but the generalization is not immediate.

To state our principal result, let $\mathbb{A}$ be the adele ring of $\mathbb{F}$ and K a maximal compact subgroup of the group $\mathrm{PGL}_2(\mathbb{A})$. The space $Y = \mathrm{PGL}_2(\mathbb{F})\backslash \mathrm{PGL}_2(\mathbb{A})/K$ is a disjoint union (indexed by a quotient of the narrow class group of $\mathbb{F}$) of finite-volume non-compact complex manifolds of dimension $[\mathbb{F} : \mathbb{Q}]$. Let μ be the quotient measure on Y induced by a fixed Haar measure on $\mathrm{PGL}_2(\mathbb{A})/K$.

Theorem 2.1.1. *Let $f : \mathrm{PGL}_2(\mathbb{A}) \to \mathbb{C}$ traverse a sequence of nondihedral holomorphic eigencuspforms of weight $(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$ as above, so that $|f|^2\, d\mu$ traverses a sequence of measures on Y. Fix a compactly supported function $\phi \in C_c(Y)$. Then*

$$\frac{\int \phi |f|^2 \, d\mu}{\int |f|^2 \, d\mu} \to \frac{\int \phi \, d\mu}{\int d\mu} \quad \text{as } \max(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty. \tag{2.1}$$

In words, the measures $|f|^2\,d\mu$ *equidistribute* as any one of the weight components k_i tend to ∞. We could normalize $d\mu$ and $|f|^2\,d\mu$ to be probability measures, in which case theorem 2.1.1 asserts that $|f|^2\,d\mu$ *converges weakly* to $d\mu$. Theorem 2.1.1 is false for certain[1] dihedral forms f that vanish identically on half of the connected components of Y; in that case, the analogous assertion that $|f|^2$ equidistributes as $\max(k_1,\dots,k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty$ on the union of the remaining connected components of Y remains true, but to simplify the exposition we shall consider only nondihedral forms in this work.

The case $\mathbb{F} = \mathbb{Q}$ of theorem 2.1.1 is the celebrated theorem of Holowinsky-Soundararajan [25], who established a quantitative rate of convergence in the limit (2.1) for a "spanning set" of functions ϕ (see §2.3). Marshall [43] proved a generalization of their result to cohomological forms over general number fields $\mathbb{F}$ that satisfy the Ramanujan conjecture, under the mild technical assumptions that $\mathbb{F}$ have narrow class number one and that the weights k_i (or the analogous archimedean parameters for fields $\mathbb{F}$ with complex places) all tend to infinity together with sufficient uniformity, precisely that $\min(k_1,\dots,k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty$ with $\min(k_1,\dots,k_{[\mathbb{F}:\mathbb{Q}]}) \geq (k_1 \cdots k_{[\mathbb{F}:\mathbb{Q}]})^\eta$ for some fixed $\eta > 0$. Since cohomological forms over totally real and imaginary quadratic number fields are known to satisfy the Ramanujan conjectures, his results are unconditional in many cases and overlap[2] with ours when $\mathbb{F}$ is totally real of narrow class number one and the weights grow uniformly in the sense just described. The essential difference between our approaches is explained in remark 4.

An important ingredient in Holowinsky's contribution to proof of theorem 2.1.1 when $\mathbb{F} = \mathbb{Q}$ is his bound

$$\sum_{n \leq x} \lambda(n)\lambda(n+l) \ll_\varepsilon \tau(l) x \log(x)^\varepsilon \prod_{p \leq x} \left(1 + \frac{\lambda(p) - 1}{p}\right)^2 \tag{2.2}$$

for any multiplicative function $\lambda : \mathbb{N} \to \mathbb{R}_{\geq 0}$ satisfying $\lambda(n) \leq \tau_m(n)$ for some positive integer m and any "shift" l satisfying $0 \neq |l| \leq x$ (see §2.3.1). A generalization of (2.2) to number fields features in Marshall's work mentioned above. We independently generalize (2.2) to number fields

[1]those induced from idele class characters on unramified totally imaginary quadratic extensions of $\mathbb{F}$; see §2.2.8.1

[2]We proved a slightly weaker form of theorem 2.1.1 in September 2009 and learned soon thereafter from Sarnak's lecture notes [54] that the overlapping results just described had been obtained earlier that year in the 2009/2010 Princeton PhD thesis of his student S. Marshall [43]. We hope that our own arguments differ sufficiently to be of interest.

that are totally real, although this restriction is not essential. The bounds that we obtain are stronger than those obtained by Holowinsky and Marshall in that we have removed the factor $\tau(l)$ appearing on the RHS of (2.2) and its generalizations (see theorem 2.4.8 and theorem 2.6.2). Although doing so is not necessary for our present purposes, this refinement has applications to the study of the distribution of mass of holomorphic forms of large level [47].

2.1.2 Plan for the Chapter

In §2.2 we introduce notation that will allow us to speak meaningfully about automorphic forms over totally real fields. In §2.3 we review the work of Holowinsky and Soundararajan over $\mathbb{F} = \mathbb{Q}$ and reduce the proof of our main result theorem 2.1.1 to that of a generalization (theorem 2.3.1) of Holowinsky's bound (1.5). The heart of our work is §2.4, in which we prove theorem 2.3.1 assuming some independent technical results that we relegate to §2.5, §2.6, §2.7 and §2.8.

2.1.3 Acknowledgements

We thank Dinakar Ramakrishnan for suggesting this problem and for his very helpful feedback and comments on earlier drafts of this chapter. We thank Fokko van de Bult for a conversation that led to a strengthening and simplification of the proof of Lemma 2.8.1. We thank Roman Holowinsky, Philippe Michel, Peter Sarnak, and K. Soundararajan for their encouragement. We thank the referee for the careful reading and comments that have helped improve our exposition.

2.2 Preliminaries

2.2.1 Number Fields

Let $\mathbb{F}$ be a totally real number field, $\mathbb{A}$ its adele ring, $\mathbb{A}_f \subset \mathbb{A}$ the subring of finite adeles, $I_{\mathbb{F}}$ the group of fractional ideals in $\mathbb{F}$, $\mathbb{F}_\infty = \mathbb{F} \otimes_{\mathbb{Q}} \mathbb{R}$, $0 \neq e_{\mathbb{F}} \in \mathrm{Hom}(\mathbb{A}/\mathbb{F}, S^1)$ the standard nontrivial additive character (i.e., normalized so that its restriction $e_{\mathbb{F}_\infty}$ to $\mathbb{F}_\infty = \mathbb{F}_\infty \times \{0\} \subset \mathbb{F}_\infty \times \mathbb{A}_f = \mathbb{A}$ is given by $e_{\mathbb{F}_\infty}(x) = e^{2\pi i \operatorname{Tr}(x)}$), $\mathbb{F}^*_{\infty+}$ the connected component of the identity in $\mathbb{F}^*_\infty$, $\mathfrak{o}$ the ring of integers in $\mathbb{F}$, $\hat{\mathfrak{o}}^* = \prod_{v<\infty} \mathfrak{o}_v^* < \mathbb{A}_f^*$ the maximal compact subgroup of the finite ideles, and $\mathfrak{o}_+^* = \mathfrak{o}^* \cap \mathbb{F}^*_{\infty+}$ the group of totally positive units of $\mathfrak{o}$, which is free abelian of rank $[\mathbb{F} : \mathbb{Q}] - 1$. Let $C_{\mathbb{F}} = \mathbb{F}^* \backslash \mathbb{A}^*$ denote the idele class group of $\mathbb{F}$ and $C_{\mathbb{F}}^1 \leq C_{\mathbb{F}}$ the (compact) kernel of the adelic absolute value.

Let $\operatorname{div} \alpha \in I_{\mathbb{F}}$ denote the fractional ideal generated by an idele $\alpha \in \mathbb{A}^*$ and $\mathrm{N}(\mathfrak{a})$ the (absolute) norm of a fractional ideal $\mathfrak{a}$. Let $\mathfrak{d}$ be the different of $\mathbb{F}$, so that $\mathfrak{d}^{-1}$ is the dual of $\mathfrak{o}$ with respect to the bilinear form $\mathbb{F} \times \mathbb{F} \ni (x, y) \mapsto e_{\mathbb{F}}(xy)$ and $\Delta_{\mathbb{F}} = \mathrm{N}(\mathfrak{d})$ is the discriminant of $\mathbb{F}$. Let $h(\mathbb{F})$ be the (finite) narrow class number of $\mathbb{F}$ and $\mathfrak{z}_1, \ldots, \mathfrak{z}_{h(\mathbb{F})}$ a set of representatives for the group of narrow ideal classes. Choose finite ideles $d_{\mathbb{F}}, z_1, z_2, \ldots, z_{[\mathbb{F}:\mathbb{Q}]} \in \mathbb{A}_f^*$ such that $\operatorname{div} d_{\mathbb{F}} = \mathfrak{d}$ and $\operatorname{div} z_j = \mathfrak{z}_j$ for $j = 1, \ldots, h(\mathbb{F})$. Then we have natural identifications

$$\mathbb{A}^* = \sqcup_{j=1}^{h(\mathbb{F})} \mathbb{F}^*(\mathbb{F}_{\infty+}^* \times z_j^{-1}\hat{\mathfrak{o}}^*), \quad \mathbb{F}^* \backslash \mathbb{A}^* / \hat{\mathfrak{o}}^* = \sqcup_{j=1}^{h(\mathbb{F})} \left((\mathbb{F}_{\infty+}^* / \mathfrak{o}_+^*) \times z_j^{-1} \right). \tag{2.3}$$

We let $\mathfrak{p}$ denote a typical prime ideal of $\mathfrak{o}$ and v a typical place of $\mathbb{F}$.

2.2.2 Asymptotic Notation

We use the asymptotic notation $\ll, \asymp$, $O()$ in the strong sense that certain inequalities should hold for all values of the parameters under consideration and not merely eventually with respect to some limit. For instance, we write $f(x,y,z) \ll_{x,y} g(x,y,z)$ 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; here $C(x,y)$ is called an *implied constant*. 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 $f(x,y,z) \ll_{x,y} g(x,y,z) \ll_{x,y} f(x,y,z)$. On the other hand, the notation $f(x) = o(g(x))$ only makes sense in the context of a limit, and we give it the standard meaning $f(x)/g(x) \to 0$.

We regard the number field $\mathbb{F}$ as fixed, so that any implied constants may depend on it without mention. We similarly regard the choice of narrow ideal class representatives $\mathfrak{z}_1, \ldots, \mathfrak{z}_{h(\mathbb{F})}$ as fixed. We let $\varepsilon \in (0, 0.01)$ denote a sufficiently small parameter and $A \geq 100$ a sufficiently large parameter, which we allow to assume finitely many distinct values throughout our analysis. We allow our implied constants to depend on ε and A without mention.

2.2.3 Real Embeddings

Set $d = [\mathbb{F} : \mathbb{Q}]$ for now. An ordering on the real embeddings $\infty_1, \ldots, \infty_d$ of $\mathbb{F}$ determines a linear inclusion $\mathbb{F} \hookrightarrow \mathbb{R}^d$ (the Minkowski embedding), which we fix. For $x \in \mathbb{R}^d$ write x_i for its ith component, so that $x_i = x^{\infty_i}$ when $x \in \mathbb{F}$. For $x, y \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}_{>0}^d$ we define $\mathbf{max}(x,y), \mathbf{min}(x,y), |x| \in \mathbb{R}^d$ and $x^\alpha \in \mathbb{R}$ by

$$\begin{aligned}
\mathbf{max}(x,y) &= (\max(x_1,y_1), \ldots, \max(x_d,y_d)), \\
\mathbf{min}(x,y) &= (\min(x_1,y_1), \ldots, \min(x_d,y_d)), \\
|x| &= (|x_1|, \ldots, |x_d|), \\
x^\alpha &= x_1^{\alpha_1} \cdots x_d^{\alpha_d}.
\end{aligned}$$

These definitions apply in particular when $x, y \in \mathbb{F} \hookrightarrow \mathbb{R}^d$. We write simply

$$\mathbf{1} = (1, \dots, 1), \quad \mathbf{0} = (0, \dots, 0),$$

so that $x^{\mathbf{1}} = x_1 \cdots x_d$ for $x \in \mathbb{R}^d$. We extend the Gamma function multiplicatively to $\mathbf{\Gamma} : (\mathbb{C} - \mathbb{Z}_{\leq 0})^d \to \mathbb{C}$ by the formula $\mathbf{\Gamma}(z) = \Gamma(z_1) \cdots \Gamma(z_d)$ for $z \in (\mathbb{C} - \mathbb{Z}_{\leq 0})^d$. As an example of our notation, for $k = (k_1, \dots, k_d) \in (2\mathbb{Z}_{\geq 1})^d$ we have

$$\frac{(4\pi\mathbf{1})^{k-\mathbf{1}}}{\mathbf{\Gamma}(k-\mathbf{1})} = \frac{(4\pi)^{k_1-1}}{\Gamma(k_1-1)} \cdots \frac{(4\pi)^{k_d-1}}{\Gamma(k_d-1)}.$$

We extend the relations $R \in \{<, \leq, \geq, >\}$ componentwise to partial orders on $\mathbb{R}^d$, writing $x \, R \, y$ to denote that $x_i \, R \, y_i$ for all $i \in \{1, \dots, d\}$; in particular, $x > \mathbf{0}$ signifies that $x_i > 0$ for all i, i.e., that x is totally positive.

2.2.4 Groups

Let $G = \mathrm{GL}(2)/\mathbb{Q}$ with the usual subgroups

$$B = \{\left(\begin{smallmatrix} * & * \\ & * \end{smallmatrix}\right)\}, \quad N = \{\left(\begin{smallmatrix} 1 & * \\ & 1 \end{smallmatrix}\right)\}, \quad A = \{\left(\begin{smallmatrix} * & \\ & * \end{smallmatrix}\right)\}, \quad Z = \{\left(\begin{smallmatrix} z & \\ & z \end{smallmatrix}\right)\}$$

and the accompanying notation

$$n(x) = \left(\begin{smallmatrix} 1 & x \\ & 1 \end{smallmatrix}\right) \in N(\mathbb{A}), \quad a(y) = \left(\begin{smallmatrix} y & \\ & 1 \end{smallmatrix}\right) \in A(\mathbb{A})$$

for $x \in \mathbb{A}$ and $y \in \mathbb{A}^*$. Put $\mathbf{X} = Z(\mathbb{A})G(\mathbb{F}) \backslash G(\mathbb{A})$.

Let $K_\infty = \mathrm{SO}(2)^{[\mathbb{F}:\mathbb{Q}]}$ be the standard maximal compact (connected) subgroup of $G(\mathbb{F}_\infty)$, let

$$K_{\text{fin}} = \prod_{v<\infty} \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\mathbb{F}_v) : a, d \in \mathfrak{o}_v, b \in \mathfrak{d}_v^{-1}, c \in \mathfrak{d}_v \right\}),$$

and let $K = K_\infty \times K_{\text{fin}}$. Then K is the conjugate by $a(1 \times d_{\mathbb{F}}^{-1})$ of the standard maximal compact subgroup of $G(\mathbb{A})$. Our choice of K_{fin} follows Shimura [62] and is convenient because the restriction to $G(\mathbb{F}_\infty)$ of a right-K_{fin}-invariant automorphic form on $G(\mathbb{A})$ has a Fourier expansion indexed by the ring of integers $\mathfrak{o}$ rather than by the inverse different $\mathfrak{d}^{-1}$.

By the Iwasawa decompositon $G(\mathbb{A}) = N(\mathbb{A})A(\mathbb{A})K$, we may define a function on $G(\mathbb{A})$ by prescribing the values it takes on elements of the form $g = n(x)a(y)kz$ with $x \in \mathbb{A}$, $y \in \mathbb{A}^*$, $k \in K$, and $z \in Z(\mathbb{A})$, provided that these values do not depend upon the choice of x, y, k, z in expressing $g = n(x)a(y)kz$.

2.2.5 Measures

We normalize Haar measures on the locally compact groups $\mathbb{A}$, $\mathbb{A}^*$, and K by requiring that

$$\mathrm{vol}(\mathbb{A}/\mathbb{F}) = \mathrm{vol}((1,e)^{[\mathbb{F}:\mathbb{Q}]} \times \hat{\mathfrak{o}}^*) = \mathrm{vol}(K) = 1.$$

We give $\mathbb{A}/\mathbb{F}$ and $C_{\mathbb{F}} = \mathbb{A}^*/\mathbb{F}^*$ the quotient measures defined with respect to the counting measures on the discrete subgroups $\mathbb{F}$, $\mathbb{F}^*$; more generally we give discrete groups such as $N(\mathbb{F}), B(\mathbb{F}), A(\mathbb{F})$, and $G(\mathbb{F})$ the counting measure and normalize accordingly the Haar measures on quotients thereof. We normalize the Haar measure on $Z(\mathbb{A})\backslash G(\mathbb{A})$ by requiring that

$$\int_{Z(\mathbb{A})B(\mathbb{Q})G(\mathbb{A})} \phi = \int_{x\in\mathbb{F}\backslash\mathbb{A}} \int_{y\in\mathbb{F}^*\backslash\mathbb{A}^*} \int_{k\in K} \phi(n(x)a(y)k)\, dx\, \frac{d^\times y}{|y|_{\mathbb{A}}}\, dk \tag{2.4}$$

for all compactly supported continuous functions ϕ on $Z(\mathbb{A})B(\mathbb{Q})\backslash G(\mathbb{A})$. This choice defines a quotient measure μ on $\mathbf{X} = Z(\mathbb{A})G(\mathbb{F})\backslash G(\mathbb{A})$. Finally, we choose a Haar measure on $C_{\mathbb{F}}^1$ so that the corresponding quotient measure on $C_{\mathbb{F}}/C_{\mathbb{F}}^1 \cong \mathbb{R}_+^*$ is the standard Haar measure $d^\times t = t^{-1}\, dt$.

2.2.6 Characters

We introduce some notation related to the Fourier transform on the idele class group $C_{\mathbb{F}} = \mathbb{F}^*\backslash\mathbb{A}^*$, and in particular its "unramified" quotient $C_{\mathbb{F}}/\hat{\mathfrak{o}}^*$.

Let $\mathfrak{X}(H)$ denote the group of (quasi-)characters on a topological abelian group H, thus $\mathfrak{X}(H)$ is the group of continuous homomorphisms $\chi : H \to \mathbb{C}^*$; a character having image in the circle group S^1 will be called a *unitary* character. For a quotient group $H'' = H/H'$ with H' closed in H, identify $\mathfrak{X}(H'')$ with the subgroup of $\mathfrak{X}(H)$ consisting of those characters having trivial restriction to H'.

Let the group $\mathfrak{X}(C_{\mathbb{F}})$ of idele class characters on $\mathbb{F}$ carry the structure of a complex manifold whose connected components are the cosets of the subgroup $\mathfrak{X}(C_{\mathbb{F}}/C_{\mathbb{F}}^1) = \{|.|^s : s \in \mathbb{C}\}$ on which the complex structure is given by s; here $|.| = |.|_{\mathbb{A}}$ is the adelic absolute value $C_{\mathbb{F}} \ni (x_v)_v \mapsto \prod |x_v|_v \in \mathbb{R}_+^*$ with $|.|_v$ the standard absolute value on the completion $\mathbb{F}_v$ of $\mathbb{F}$, so that multiplication by x_v scales the Haar measure on $\mathbb{F}_v$ by $|x_v|_v$.

Since $C_{\mathbb{F}}^1$ is compact, for each $\chi \in \mathfrak{X}(C_{\mathbb{F}})$ we have $|\chi| = |.|^\sigma$ for some $\sigma \in \mathbb{R}$, which we call the *real part* of χ and denote by $\sigma = \mathrm{Re}(\chi)$. Let $\mathfrak{X}(C_{\mathbb{F}})(c)$ denote the set of idele class characters having real part c.

Let

$$\mathfrak{X}(C_{\mathbb{F}})[2] := \{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}) : \chi_0^2 = 1\}$$

denote the group of *quadratic* idele class characters. This is not to be confused with the set

$\mathfrak{X}(C_{\mathbb{F}})(2)$ of idele class characters χ having real part $\mathrm{Re}(\chi) = 2$.

Let $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$ denote the restriction of an idele class character $\chi \in \mathfrak{X}(C_{\mathbb{F}})$ to $\mathbb{F}_\infty^*$. Then χ_∞ is of the form

$$y \mapsto \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} \mathrm{sgn}(y_j)^{\varepsilon_j} |y_j|^{ir_j} \quad \text{if } y = (y_1, \ldots, y_{[\mathbb{F}:\mathbb{Q}]}) \in (\mathbb{R}^{[\mathbb{F}:\mathbb{Q}]})^* = \mathbb{F}_\infty^* \tag{2.5}$$

for some $\varepsilon_j \in \{0, 1\}$ and $r_j \in \mathbb{C}$; the character χ_∞ is unitary if and only if each $r_j \in \mathbb{R}$. For a place v of $\mathbb{F}$, let χ_v be the restriction of χ to $\mathbb{F}_v^* \hookrightarrow \mathbb{A}^*$; in particular, $\chi_{\infty_j} = [y_j \mapsto \mathrm{sgn}(y_j)^{\varepsilon_j}|y_j|^{ir_j}]$ is the restriction of χ_∞ as above to the jth factor of $(\mathbb{R}^{[\mathbb{F}:\mathbb{Q}]})^*$,

The group $\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ of unramified idele class characters χ is a subgroup of the group $\mathfrak{X}(C_{\mathbb{F}})$ of all idele class characters; here and elsewhere unramified means "unramified at all finite places." Set $\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(c) := \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*) \cap \mathfrak{X}(C_{\mathbb{F}})(c)$ for any $c \in \mathbb{R}$ and $\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2] := \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*) \cap \mathfrak{X}(C_{\mathbb{F}})[2]$.

Let

$$\xi_{\mathbb{F}} : \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*) \to \mathbb{P}^1(\mathbb{C})$$

be the (completed) Dedekind zeta function, defined for unramified idele class characters of real part $\mathrm{Re}(\chi) > 1$ by the Euler product $\xi_{\mathbb{F}}(\chi) = \prod_v \zeta_v(\chi_v)$ and in general by meromorphic continuation, where $\zeta_{\mathfrak{p}}(v) = (1 - \chi_{\mathfrak{p}}(\varpi_{\mathfrak{p}}))^{-1}$ for $\varpi_{\mathfrak{p}}$ a generator of $\mathfrak{p} \subset \mathbb{F}_{\mathfrak{p}}$ and $\zeta_{\infty_j}(\chi_{\infty_j}) = \Gamma_{\mathbb{R}}(ir_j + \varepsilon_j)$ if χ_∞ is given by (2.5); here $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$. For $s \in \mathbb{C}$ let $\xi_{\mathbb{F}}(s) := \xi_{\mathbb{F}}(|.|^s)$, which agrees with the usual definition. Hecke proved that $\xi_{\mathbb{F}}$ is holomorphic away from its simple pole at $\chi = |.|$ and satisfies a functional equation relating its values at χ and $|.|\chi^{-1}$.

Let $\Psi \in C_c^\infty(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ be a test function. For each character $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ let $\Psi^\wedge(\chi)$ be the Fourier-Mellin transform of Ψ at χ normalized so that the inversion formula

$$\Psi(y) = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(c)} \Psi^\wedge(\chi)\chi(y)\,\frac{d\chi}{2\pi i} \tag{2.6}$$

holds, where $\int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(c)}$ denotes the contour integral over unramified idele class characters χ having real part $c > 1$ taken in the usual vertical sense, precisely

$$\int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(c)} \Psi^\wedge(\chi)\chi(y)\,\frac{d\chi}{2\pi i} := \sum_{\chi_0 \in \frac{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)}{\mathfrak{X}(C_{\mathbb{F}}/C_{\mathbb{F}}^1)}} \int_{(c)} \Psi^\wedge(\chi_0|.|^s)\chi_0(y)|y|_{\mathbb{A}}^s\,\frac{ds}{2\pi i},$$

where $\int_{(c)}$ denotes the vertical contour integral taken over $\mathrm{Re}(s) = c$ from $c - i\infty$ to $c + i\infty$, and as representatives for the quotient $\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)/\mathfrak{X}(C_{\mathbb{F}}/C_{\mathbb{F}}^1)$ one may take the image of the discrete group $\mathfrak{X}(C_{\mathbb{F}}^1/\hat{\mathfrak{o}}^*)$ under pullback by a section of the inclusion $C_{\mathbb{F}}^1 \hookrightarrow C_{\mathbb{F}}$. By our normalization of measures (see §2.2.5), the forward transform is given explicitly by

$$\Psi^\wedge(\chi) = \frac{1}{\mathrm{vol}(C_{\mathbb{F}}^1)} \int_{C_{\mathbb{F}}} \Psi(y)\chi^{-1}(y)\, d^\times y. \tag{2.7}$$

The analytic conductor [32] of an unramified idele class character $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ having archimedean component (2.5) is defined to be

$$C(\chi) = \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} (3 + |r_j|); \tag{2.8}$$

the number 3 is unimportant and present only so that $\log C(\chi)$ is never too small. Repeated "partial integration" shows that $\Psi^\wedge(\chi) \ll_{\Psi,A} C(\chi)^{-A}$ for any test function $\Psi \in C_c^\infty(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ and any positive integer A, uniformly for $\mathrm{Re}(\chi)$ in any bounded set. Concretely, we have natural short exact sequences

$$1 \to \mathbb{F}_{\infty+}^*/\mathfrak{o}_+^* \to C_{\mathbb{F}}/\hat{\mathfrak{o}}^* \to \mathrm{Cl}_{\mathbb{F}}^+ \to 1,$$

and

$$1 \to \mathbb{F}_{\infty+}^1/\mathfrak{o}_+^* \to \mathbb{F}_{\infty+}^*/\mathfrak{o}_+^* \xrightarrow{x \mapsto x^1} \mathbb{R}_+^* \to 1,$$

where $\mathrm{Cl}_{\mathbb{F}}^+ = C_{\mathbb{F}}/(\mathbb{F}_{\infty+}^* \times \hat{\mathfrak{o}}^*)$ is the (finite) narrow class group of $\mathbb{F}$ and $\mathbb{F}_{\infty+}^1$ is the subgroup $\{(x_i) : \prod x_i = 1\}$ of $\mathbb{F}_{\infty+}^*$. Thus $C_{\mathbb{F}}/\hat{\mathfrak{o}}^*$ is an extension of a finite group by an extension of $\mathbb{R}_+^*$ by a compact torus, so the assertion $\Psi^\wedge(\chi) \ll_{\Psi,A} C(\chi)^{-A}$ reduces to the familiar decay properties of the Fourier transform of a test function on a finite product of Euclidean lines and circles.

2.2.7 Fourier Expansions

Suppose that $\phi : \mathbf{X} \to \mathbb{C}$ is continuous and right-K-invariant. By the Iwasawa decomposition, ϕ is determined by the values $\phi(n(x)a(y))$ for $x \in \mathbb{A}, y \in \mathbb{A}^*$. If ϕ is assumed merely to be right-K_{fin}-invariant but transforms under a unitary character of K_∞, then $|\phi|^2$ is still determined by the values $\phi(n(x)a(y))$. In either case, the left-$B(\mathbb{F})$-invariance of ϕ implies a Fourier expansion

$$\phi(n(x)a(y)) = \phi_0(y) + \sum_{n \in \mathbb{F}^*} \kappa_\phi(ny) e_{\mathbb{F}}(nx) \tag{2.9}$$

for some functions ϕ_0 on $C_{\mathbb{F}}/\hat{\mathfrak{o}}^* = \mathbb{F}^* \backslash \mathbb{A}^*/\hat{\mathfrak{o}}^*$ and κ_ϕ on $\mathbb{A}^*/\hat{\mathfrak{o}}^*$ (see [71]).

We say that the Fourier expansion (2.9) of ϕ is *factorizable* if for each $y \times z \in \mathbb{F}_\infty^* \times \mathbb{A}_f^* = \mathbb{A}^*$ we have

$$\kappa_\phi(y \times z) = \kappa_{\phi,\infty}(y) \frac{\lambda_\phi(\mathrm{div}\, z)}{\mathrm{N}(\mathrm{div}\, z)^{1/2}}, \tag{2.10}$$

where $\lambda_\phi : I_{\mathbb{F}} \to \mathbb{C}$ is a weakly multiplicative function supported on the monoid of integral ideals and $\kappa_{\phi,\infty}(y) = \prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} \kappa_{\phi,\infty_j}(y_j)$ for some functions $\kappa_{\phi,\infty_j} : \mathbb{R}^* \to \mathbb{C}$.

2.2.8 Automorphic Forms

We shall consider various kinds of automorphic forms throughout this chapter. In this section we give them convenient names and state their relevant properties.

2.2.8.1 Holomorphic eigencuspforms

By a *holomorphic eigencuspform* $f : \mathbf{X} \to \mathbb{C}$ of weight $k = (k_1, \dots, k_{[\mathbb{F}:\mathbb{Q}]})$ (here and always each k_j is a *positive* even integer, for simplicity) we mean an arithmetically normalized cuspidal holomorphic Hilbert modular form of weight k, full level, and trivial central character, that is furthermore an eigenfunction of the algebra of Hecke operators. Precise definitions in both the classical and adelic languages appear in Shimura's paper [62]; for our purposes, it is necessary to know only that f is right K_{fin}-invariant, transforms under a (specific) unitary character of K_∞, and has a factorizable Fourier expansion (2.9) with $f_0 \equiv 0$ and

$$\kappa_{f,\infty_j}(y) = \begin{cases} y^{k_j/2} e^{-2\pi y} & \text{for } y > 0, \\ 0 & \text{for } y < 0 \end{cases} \tag{2.11}$$

for each infinite place ∞_j of $\mathbb{F}$. The "Ramanujan bound" for f [2] asserts[3] that $|\lambda_f(\mathfrak{a})| \leq \tau(\mathfrak{a})$ for each integral ideal $\mathfrak{a}$, where τ is the divisor function (multiplicative, $\mathfrak{p}^k \mapsto k+1$); this improves an earlier result of Brylinski-Labesse, which asserts that $|\lambda_f(\mathfrak{p})| \leq 2$ for a full density set of primes $\mathfrak{p}$.

To f and an unramified idele class character $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ of sufficiently large real part we associate the finite part of the adjoint L-function

$$L(\operatorname{ad} f, \chi) = \prod_{\mathfrak{p}} L_{\mathfrak{p}}(\operatorname{ad} f, \chi)$$

and its completion $\Lambda(\operatorname{ad} f, \chi) = L_\infty(\operatorname{ad} f, \chi) L(\operatorname{ad} f, \chi) = \prod_v L_v(\operatorname{ad} f, \chi)$, where the local factors are as in [70, §3.1.1]. It is known [61, 13] that $\chi \mapsto L(\operatorname{ad} f, \chi)$ continues meromorphically to a function on $\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ whose only possible poles are simple and at $\chi = \chi_0|.|$ for $\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]$ a quadratic character. Call f *nondihedral* if $L(\operatorname{ad} f, \cdot) : \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*) \to \mathbb{P}^1(\mathbb{C})$ is entire; this is known to be the case precisely when f is not induced from an idele class character of a

[3] the parity conditions on the weight of f are satisfied because f has trivial central character, hence the k_i are all even

quadratic extension of $\mathbb{F}$ [13, 39]. Note that unlike when $\mathbb{F} = \mathbb{Q}$ or $h(\mathbb{F}) = 1$, in general (e.g., for $\mathbb{F} = \mathbb{Q}(\sqrt{3})$) there may exist dihedral cusp forms of full level and trivial central character, which we shall exclude from our analysis.

2.2.8.2 Maass eigencuspforms

By a *Maass eigencuspform* $\phi : \mathbf{X} \to \mathbb{C}$ of Laplace eigenvalue $(\frac{1}{4} + r_1^2, \ldots, \frac{1}{4} + r_{[\mathbb{F}:\mathbb{Q}]}^2) \in \mathbb{R}_{>0}^{[\mathbb{F}:\mathbb{Q}]}$ and parity $(\varepsilon_1, \ldots, \varepsilon_{[\mathbb{F}:\mathbb{Q}]}) \in \{0,1\}^{[\mathbb{F}:\mathbb{Q}]}$ we mean an arithmetically normalized Hilbert-Maass cusp form on $\mathbf{X}$ of given Laplace eigenvalues and parity, full level and trivial central character, that is furthermore an eigenfunction of the algebra of Hecke operators. For our purposes this means that ϕ is right-K-invariant and has a factorizable Fourier expansion (2.9) with $\phi_0 \equiv 0$ and

$$\kappa_{\phi,\infty_j}(y) = 2|y|^{1/2} K_{ir_j}(2\pi|y|)\,\mathrm{sgn}(y)^{\varepsilon_j} \tag{2.12}$$

for each infinite place ∞_j and all $y \in \mathbb{R}^*$; here K_{ir_j} is the modified Bessel function of the second kind. The trivial "Hecke bound" asserts that $\lambda_\phi(\mathfrak{a}) \leq \tau(\mathfrak{a})\,\mathrm{N}(\mathfrak{a})^{1/2}$. The "Rankin-Selberg bound," also known as the "Ramanujan bound on average," asserts that

$$\sum_{\mathrm{N}(\mathfrak{a}) \leq x} |\lambda_\phi(\mathfrak{a})|^2 \ll_\phi x \tag{2.13}$$

and follows as in [29, §8.2] from the analytic properties of the Rankin-Selberg L-series attached to $\phi \times \phi$ [33].

2.2.8.3 Eisenstein series

Let $\chi \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)$ be an unramified idele class character. Writing $y(g) = y$ for $g = n(x)a(y)kz$, the map $B(\mathbb{F})\backslash G(\mathbb{A}) \ni g \mapsto \chi(y(g))$ is well-defined. The *Eisenstein series*

$$E(\chi, g) = \sum_{\gamma \in B(\mathbb{F})\backslash G(\mathbb{F})} \chi(y(\gamma g)) \tag{2.14}$$

converges normally in g and uniformly in χ for $\mathrm{Re}(\chi) \geq 1 + \delta > 0$, and continues meromorphically to the union of half-planes on which $\mathrm{Re}(\chi) \geq \frac{1}{2}$, where $\chi \mapsto E(\chi, \cdot)$ is holomorphic with the exception of simple poles at $\chi = |.|\chi_0$ of locally constant residue proportional to $g \mapsto \chi_0(\det(g))$ for each unramified *quadratic* idele class character $\chi_0 \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)[2]$ (see [14]). The functions $E(\chi, \cdot) : g \mapsto E(\chi, g)$ descend to $\mathbf{X} = Z(\mathbb{A})G(\mathbb{F})\backslash G(\mathbb{A})$ and are right-K-invariant by construction.

The scaled Eisenstein series $\phi = \Delta_\mathbb{F}^{-1}\chi(d_\mathbb{F})^{-2}\xi_\mathbb{F}(\chi^2)E(\chi, \cdot)$ admits a factorizable Fourier expansion (2.9) with

$$\phi_0(y) = \Delta_{\mathbb{F}}^{-1}\chi(d_{\mathbb{F}})^{-2}\xi_{\mathbb{F}}(\chi^2)\chi(y) + \Delta_{\mathbb{F}}^{-1/2}\xi_{\mathbb{F}}(\chi^2|.|^{-1})\chi^{-1}(y)|y|, \tag{2.15}$$

$$\kappa_\phi(y \times z) = \kappa_{(\chi|.|^{-1/2})_\infty}(y)\frac{\lambda_{(\chi|.|^{-1/2})}(\operatorname{div} z)}{\mathrm{N}(\operatorname{div} z)^{1/2}}$$

as in §2.2.7, where for $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ with χ_∞ given by (2.5), we set

$$\kappa_{\chi_{\infty_j}}(y) = 2|y|^{1/2}K_{ir_j}(2\pi|y|)\operatorname{sgn}(y)^{\varepsilon_j}, \quad \lambda_\chi(\mathfrak{p}^k) = \sum_{i=0}^{k}\chi(\mathfrak{p})^i\chi^{-1}(\mathfrak{p})^{k-i}; \tag{2.16}$$

for a convenient tabulation of such Fourier expansions of Eisenstein series see [3].

If $\chi|.|^{-1/2}$ is a unitary character (equivalently, $\mathrm{Re}(\chi) = \frac{1}{2}$, i.e., $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(\frac{1}{2})$), call $E(\chi, g)$ a *unitary Eisenstein series*; in that case $|\lambda_{\chi|.|^{-1/2}}(\mathfrak{a})| \leq \tau(\mathfrak{a})$.

2.2.8.4 Incomplete Eisenstein series

To a test function $\Psi \in C_c^\infty(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ attach the *incomplete Eisenstein series* $E(\Psi, \cdot) : \mathbf{X} \to \mathbb{C}$ by the formula

$$E(\Psi, g) = \sum_{\gamma \in B(\mathbb{F})\backslash G(\mathbb{F})} \Psi(y(\gamma g)) \tag{2.17}$$

with $y(\gamma g)$ as in §2.2.8.3. Write $\phi = E(\Psi, \cdot)$. We have $\Psi^\wedge(|.|)\operatorname{res}_{s=1} E(|.|^s, \cdot) = \mu(\phi)/\mu(1)$ (see §2.3.3), so by shifting the contour in the integral representation $E(\Psi, \cdot) = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(2)} \Psi^\wedge(\chi)E(\chi, \cdot)\,\frac{d\chi}{2\pi i}$ to the union of lines $\mathrm{Re}(\chi) = \frac{1}{2}$ (see [14] and [29, §7.3]), we obtain

$$\begin{aligned} E(\Psi, g) = \frac{\mu(\phi)}{\mu(1)} &+ \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0)\chi_0(\det g) \\ &+ \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi)E(\chi, g)\,\frac{d\chi}{2\pi i} \end{aligned} \tag{2.18}$$

for some constants $c_\Psi(\chi_0) = \mu(1)^{-1}\int_{\mathbf{X}} E(\Psi, \cdot)(\chi_0 \circ \det)$ whose precise values are not important for our purposes. Taking the Fourier expansions of both sides gives

$$\phi_0(y) = \frac{\mu(\phi)}{\mu(1)} + \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0)\chi_0(y) + O_\phi(|y|^{1/2}), \tag{2.19}$$

$$\kappa_\phi(y \times z) = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)} \frac{\Psi^\wedge(|.|^{1/2}\chi)}{\xi_{\mathbb{F}}(|.|\chi^2)\chi(d_{\mathbb{F}})^{-2}}\kappa_{\chi,\infty}(y)\frac{\lambda_\chi(\operatorname{div} z)}{\mathrm{N}(\operatorname{div} z)^{1/2}}\,\frac{d\chi}{2\pi i}. \tag{2.20}$$

2.2.9 Masses

Recall the measure μ defined on the space $\mathbf{X} = Z(\mathbb{A})G(\mathbb{F})\backslash G(\mathbb{A})$ in §2.2.5. For $\phi \in L^1(\mathbf{X}, \mu)$ let $\mu(\phi) = \int_{\mathbf{X}} \phi \, d\mu$. To our varying nondihedral holomorphic eigencuspform f we associate the finite measure $d\mu_f = |f|^2 \, d\mu$ and write accordingly $\mu_f(\phi) = \int_{\mathbf{X}} \phi |f|^2 \, d\mu$. In particular, writing 1 for the constant function on $\mathbf{X}$, we see that $\mu(1)$ is the volume of $\mathbf{X}$ and $\mu_f(1)$ the mass of f, i.e., its squared norm in $L^2(\mathbf{X}, \mu)$. With this notation, the conclusion of theorem 2.1.1 is that for any compactly supported, continuous, right-K-invariant function ϕ on $\mathbf{X}$, we have

$$\frac{\mu_f(\phi)}{\mu_f(1)} \to \frac{\mu(\phi)}{\mu(1)}$$

as any of the weight components of f tend to ∞. It suffices to show this for ϕ a Maass eigencuspform or incomplete Eisenstein series as in §2.2.8.2 and §2.2.8.4.

The special value $L(\mathrm{ad}\, f, 1)$ enters our analysis through the Rankin-type formula

$$\mu_f(1) = \frac{\mathbf{\Gamma}(k)}{c_1(\mathbb{F})(4\pi\mathbf{1})^{k-\mathbf{1}}} L(\mathrm{ad}\, f, 1), \quad c_1(\mathbb{F}) := \frac{(4\pi^2)^{[\mathbb{F}:\mathbb{Q}]}}{2\Delta_{\mathbb{F}}^{3/2}}. \tag{2.21}$$

We sketch the standard calculation. Recall the measure normalization (2.4) and the choice of compact subgroup K (§2.2.4) on which we base our definition (§2.2.8.3) of $E(s, \cdot)$. For $\mathrm{Re}(s) > 1$ we find by unfolding that

$$\begin{aligned}
\mu_f(E(s,\cdot)) &= \int_{Z(\mathbb{A})B(\mathbb{F})\backslash G(\mathbb{A})} |y(g)|_{\mathbb{A}}^{s} |f|^2(g)\, dg \\
&= \int_{x \in \mathbb{F}\backslash\mathbb{A}} \int_{y \in \mathbb{F}^*\backslash\mathbb{A}^*} |y|_{\mathbb{A}}^{s-1} |f|^2(n(x)a(y))\, dx\, d^\times y \\
&= \prod_v \int_{y \in \mathbb{Q}_v^*} |y|_v^{s-1} |\kappa_f(y)|^2\, d^\times y \\
&= \Lambda(\mathrm{ad}\, f, s) \frac{\xi_{\mathbb{F}}(s)}{\xi_{\mathbb{F}}(2s)} \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} 2^{-k_i - 1}
\end{aligned}$$

by local calculations as conveniently tabulated in [70, §3.2.1]. Since the Fourier expansion (2.15) implies

$$\mathrm{res}_{s=1} E(s, \cdot) = \Delta_{\mathbb{F}}^{-3/2} \frac{\mathrm{res}_{s=1} \xi_{\mathbb{F}}(s)}{2\xi_{\mathbb{F}}(2)}$$

and by definition [70, §3.1.1]

$$L_\infty(\mathrm{ad}\, f, 1) \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} 2^{-k_i - 1} = (4\pi^2)^{-[\mathbb{F}:\mathbb{Q}]} \frac{\mathbf{\Gamma}(k)}{(4\pi\mathbf{1})^{k-\mathbf{1}}},$$

we obtain the claimed formula (2.21).

2.3 Brief Review of Holowinsky-Soundararajan

In this section we summarize the Holowinsky-Soundararajan [25] proof of theorem 2.1.1 when $\mathbb{F} = \mathbb{Q}$ and indicate which of their arguments require generalization when $\mathbb{F}$ is a general totally real number field. Their proof combines

(1) the independent arguments of Holowinsky [24], and

(2) the independent arguments of Soundararajan [66],

(3) the joint Holowinsky-Soundararajan synthesis of (1) and (2).

As we shall see, Soundararajan's independent arguments and the Holowinsky-Soundararajan synthesis generalize painlessly, so the essential difficulty is to generalize Holowinsky's arguments. In this section, f is a holomorphic eigencuspform of weight $k = (k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$. Recall from §2.2.3 that $k^{\mathbf{1}} := k_1 \ldots k_{[\mathbb{F}:\mathbb{Q}]}$, thus when $\mathbb{F} = \mathbb{Q}$ we have $k = (k_1)$ and $k^{\mathbf{1}} = k_1$.

2.3.1 Holowinsky's Independent Arguments

We begin by simultaneously recalling Holowinsky's main result [24, Cor 3] and stating our generalization thereof. Define for each holomorphic eigencuspform f and each real number $x \geq 2$ the quantities

$$M_f(x) = \frac{\log(x)^{-2}}{L(\operatorname{ad} f, 1)} \prod_{\mathrm{N}(\mathfrak{p}) \leq x} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right), \tag{2.22}$$

$$R_f(x) = \frac{x^{-1/2}}{L(\operatorname{ad} f, 1)} \sum_{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} \int_{(1/2)} \left| \frac{L(\operatorname{ad} f, \chi_0|.|^s)}{C(\chi_0|.|^s)^{10}} \right| |ds|. \tag{2.23}$$

Here $C(\chi_0|.|^s) \asymp |s|^{[\mathbb{F}:\mathbb{Q}]}$ since χ_0 is quadratic.

Theorem 2.3.1. *Let f be a nondihedral holomorphic eigencuspform of weight $k = (k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$. If ϕ is a Maass eigencuspform, then*

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

If ϕ is an incomplete Eisenstein series, then

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

We prove theorem 2.3.1 in §2.4 by combining the independent results of §2.5, §2.8 and §2.6; doing so is our main task in this work. Holowinsky [24, Cor 3] established the case $\mathbb{F} = \mathbb{Q}$ of theorem 2.3.1, in which the "nondihedral" hypothesis is vacuously satisfied. We briefly recall his argument. Take $\mathbb{F} = \mathbb{Q}$ and denote by k the weight of f. Suppose for simplicity that ϕ is a Maass eigencuspform. Holowinsky defines for a fixed test function $h \in C_c^\infty(\mathbb{R}_+^*)$ the integral

$$S_l(Y) = \int_{y \in \mathbb{R}_+^*} h(Yy) \int_{x \in \mathbb{R}/\mathbb{Z}} (\phi_l |f|^2)(x+iy) \, \frac{dx\, dy}{y^2},$$

where $\phi(z) = \sum_l \phi_l(z)$ with $\phi_l(z+\xi) = e^{2\pi i l \xi}\phi_l(z)$ for $\xi \in \mathbb{R}$, and establishes [24, Theorem 1] for any $Y \geq 1$ and $\varepsilon > 0$ the asymptotic formula

$$\frac{\int \phi |f|^2}{\int |f|^2} = cY^{-1} \sum_{0<|l|<Y^{1+\varepsilon}} S_l(Y) + O_{\phi,\varepsilon}(Y^{-1/2}) \tag{2.24}$$

where c is an explicit nonzero constant depending only upon the test function h; he shows moreover that

$$\frac{S_l(Y)}{Y} \ll_{\phi,\varepsilon} \frac{|\phi_l(a(Y^{-1}))|}{L(\mathrm{ad}\, f, 1)} \left[\frac{1}{Yk} \sum_{\substack{n \in \mathbb{N} \\ m := n+l \in \mathbb{N}}} |\lambda_f(m)\lambda_f(n)| h\left(\frac{Y\left(\frac{k-1}{4\pi}\right)}{\frac{m+n}{2}} \right) + \frac{(Yk)^\varepsilon}{k} \right]. \tag{2.25}$$

He then proves [24, Theorem 2] (in somewhat greater generality) that for each $\varepsilon \in (0,1)$, each $x \gg_\varepsilon 1$, and each $l \in \mathbb{Z}$ for which $0 \neq |l| \leq x$, we have

$$\sum_{n \leq x} |\lambda_f(m)\lambda_f(n)| \ll \tau(l) \frac{x}{\log(x)^{2-\varepsilon}} \prod_{p \leq X} \left(1 + \frac{2|\lambda_f(p)|}{p}\right). \tag{2.26}$$

From this he deduces the cuspidal case of theorem 2.3.1 for $\mathbb{F} = \mathbb{Q}$. We generalize and refine (2.24), (2.25) and (2.26) in §2.5, §2.8 and §2.6, respectively; among other refinements, we show that (a generalization to totally real number fields of) the bound (2.26) holds without the factor $\tau(l)$. The main complication is the manner in which these ingredients fit together to yield theorem 2.3.1 when $\mathbb{F} \neq \mathbb{Q}$; this is the crux of our argument, which we present in §2.4. Specifically, recall that for a totally real number field $\mathbb{F}$ of degree $d = [\mathbb{F} : \mathbb{Q}]$, our naïve generalization of (2.24) and (2.25) leaves us with the task of showing that a sum of roughly $x \log(x)^{d-1}$ terms is small relative to x (with x a bit larger than $k^{\mathbf{1}}$), which seems beyond the limits of any method that does not exploit cancellation in the sum of $\lambda_f(m)\lambda_f(n)$. By discarding a large number of these terms trivially through a refinement of (2.25), we reduce to the more tractable problem of showing that a sum of roughly $x \log(x)^\varepsilon$ terms is small relative to x.

2.3.2 Soundararajan's Independent Arguments

Let ϕ be a Maass eigencuspform, and suppose that $\mathbb{F} = \mathbb{Q}$. Watson's formula [70, Theorem 3] asserts that

$$\left|\frac{\mu_f(\phi)}{\mu_f(1)}\right|^2 = c(\mathbb{F}, \phi)\frac{\Lambda(\phi \times f \times f, \frac{1}{2})}{\Lambda(\operatorname{ad} f, 1)^2} \tag{2.27}$$

where $c(\mathbb{Q}, \phi) = \mu(|\phi|^2)/8\Lambda(\operatorname{ad}\phi, 1)$ is a nonzero constant unimportant for our purposes and $\Lambda(\cdots, s)$ is the completed L-function for $L(\cdots, s)$ with local factors as in [70, §3.1.1]. The identity (2.27) with $c(\mathbb{F}, \phi) \neq 0$ holds for totally real $\mathbb{F}$ by Ichino's general triple product formula [26] together with Watson's calculations of the local zeta integrals of Harris-Kudla [19] at the real places. When $\mathbb{F} = \mathbb{Q}$, Soundararajan [66, Ex 2] proves that

$$L(\phi \times f \times f, \tfrac{1}{2}) \ll_{\phi,\varepsilon} \frac{k^{\mathbf{1}}}{\log(k^{\mathbf{1}})^{1-\varepsilon}}. \tag{2.28}$$

His argument applies verbatim when $\mathbb{F}$ is totally real: it relies only upon the Ramanujan bound for the local components of f and the Rankin-Selberg theory for $\phi \times \phi$, noting that the analytic conductor of $\phi \times f \times f$ is $\asymp_\phi (k^{\mathbf{1}})^4$. By Stirling's formula as in the $\mathbb{F} = \mathbb{Q}$ case, we obtain

$$\frac{\int \phi |f|^2}{\int |f|^2} \ll_{\phi,\varepsilon} \frac{\log(k^{\mathbf{1}})^{-1/2+\varepsilon}}{L(\operatorname{ad} f, 1)}. \tag{2.29}$$

Now let $\phi = E(\chi, \cdot)$ be the unitary Eisenstein series associated as in §2.2.8.3 to an unramified idele class character $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(\frac{1}{2})$ of real part $\frac{1}{2}$, and suppose that $\mathbb{F} = \mathbb{Q}$. (Since $C_{\mathbb{Q}}/\hat{\mathbb{Z}}^* \cong \mathbb{R}_+^*$, we have $\chi = |.|^{1/2+it}$ for some $t \in \mathbb{R}$.) Soundararajan [66, p7] shows by the unfolding method, Stirling's formula and his weak subconvex bounds for $L(\operatorname{ad} f, \chi)$ [66, Ex 1], the last of which makes use of the known Ramanujan bound for f, that

$$\frac{\mu_f(\phi)}{\mu_f(1)} \ll_\varepsilon C(\chi)^2 \frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}, \tag{2.30}$$

and [66, p2]

$$|L(\operatorname{ad} f, \chi)| \ll_\varepsilon \frac{(k^{\mathbf{1}})^{1/2} C(\chi)^{3/4}}{\log(k^{\mathbf{1}})^{1-\varepsilon}}. \tag{2.31}$$

By the modularity of $L(\operatorname{ad} f, \chi)$ as the L-function of an automorphic form on $GL(3)$ [13], its Rankin-Selberg theory, and the lower bound

$$L(\operatorname{ad} f, 1) \gg \log(k^{\mathbf{1}})^{-1} \tag{2.32}$$

due to Hoffstein-Lockhart-Goldfeld-Hoffstein-Lieman [22] (which is available for general $\mathbb{F}$, see

[3, §2.9]), Soundararajan deduces [25, Lem 1] in his joint paper with Holowinsky that

$$R_f(k^{\mathbf{1}}) \ll_\varepsilon \frac{\log(k^{\mathbf{1}})^\varepsilon}{\log(k^{\mathbf{1}}) L(\operatorname{ad} f, 1)} \ll \log(k^{\mathbf{1}})^\varepsilon. \tag{2.33}$$

The same argument establishes (2.30), (2.31), (2.33) for general totally real number fields $\mathbb{F}$.

2.3.3 The Holowinsky-Soundararajan Synthesis

In their joint work [25], Holowinsky and Soundararajan show [24, Lem 3] for $\mathbb{F} = \mathbb{Q}$ that

$$M_k(f) \ll \log(k^{\mathbf{1}})^{1/6} \log\log(k^{\mathbf{1}})^{9/2} L(\operatorname{ad} f, 1)^{1/2}, \tag{2.34}$$

and their proof applies for general $\mathbb{F}$. Subsituting the bound (2.34) into theorem 2.3.1 and combining with Soundararajan's estimate (2.29) yields for each Maass eigencuspform ϕ that

$$\frac{\mu_f(\phi)}{\mu_f(1)} \ll_{\phi,\varepsilon} \min\left(\frac{\log(k^{\mathbf{1}})^{-1/2+\varepsilon}}{L(\operatorname{ad} f, 1)}, \log(k^{\mathbf{1}})^{1/12+\varepsilon} L(\operatorname{ad} f, 1)^{1/4} \right). \tag{2.35}$$

It follows as in [25, Proof of Thm 1] that $\mu_f(\phi)/\mu_f(1) \ll_{\phi,\varepsilon} \log(k^{\mathbf{1}})^{-1/30+\varepsilon} = o(1)$, and the same argument applies in the totally real case as soon as one has established theorem 2.3.1.

Holowinsky and Soundararajan show [25, p10] that Soundararajan's bound (2.30) for unitary Eisenstein series also applies to incomplete Eisenstein series via the Mellin inversion formula. Specifically, they show for $\mathbb{F} = \mathbb{Q}$ and $\phi = E(\Psi, \cdot)$ that

$$\left| \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} \right| \ll_{\phi,\varepsilon} \frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}. \tag{2.36}$$

Their argument generalizes to the totally real case by replacing the Mellin inversion on $\mathbb{R}_+^* \cong C_{\mathbb{Q}}/\hat{\mathbb{Z}}^*$ with that on $C_{\mathbb{F}}/\hat{\mathfrak{o}}^*$, as we now describe. Let $\Psi \in C_c^\infty(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ and $\phi = E(\Psi, \cdot)$. By the Mellin formula (see §2.2.6)

$$\phi = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(2)} \Psi^\wedge(\chi) E(\chi, \cdot) \, \frac{d\chi}{2\pi i}$$

and the meromorphic nature of $E(\chi, \cdot)$ (see §2.2.8.3 or [14]), we have

$$\begin{aligned} \mu_f(\phi) = &\sum_{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} \Psi^\wedge(\chi_0) \operatorname{res}_{s=1} \mu_f(E(\chi_0 |.|^s, \cdot)) \\ &+ \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi) \mu_f(E(\chi, \cdot)) \, \frac{d\chi}{2\pi i}, \end{aligned} \tag{2.37}$$

where the interchanges here and those that follow are justified by absolute convergence owing to the rapid decay of f and Ψ and the moderate growth of $E(\chi, \cdot)$. By the unfolding method as in §2.2.9, the residue $\operatorname{res}_{s=1} \mu_f(E(\chi_0|.|^s, \cdot))$ coincides with $\operatorname{res}_{s=1} \Lambda(\operatorname{ad} f, \chi_0|.|^s)\xi_{\mathbb{F}}(\chi_0|.|^s)$ up to a nonzero scalar. Suppose now that f is nondihedral in the sense of §2.2.8.1, so that $s \mapsto \Lambda(\operatorname{ad} f, \chi_0|.|^s)$ is entire. Then since $\xi_{\mathbb{F}}$ is holomorphic away from its pole at $\chi = |.|$, we see that $\operatorname{res}_{s=1} \mu_f(E(\chi_0|.|^s, \cdot)) = 0$ if $\chi_0 \neq 1$. If $\chi_0 = 1$, then

$$\Psi^\wedge(|.|) \operatorname{res}_{s=1} \mu_f(E(|.|^s, \cdot)) = \mu_f(1)\Psi^\wedge(|.|) \operatorname{res}_{s=1} E(|.|^s, \cdot).$$

We have $\Psi^\wedge(|.|) \operatorname{res}_{s=1} E(|.|^s, \cdot) = \mu(\phi)/\mu(1)$ because both sides are equal to the coefficient of the constant function 1 in the spectral decomposition of $\phi \in L^2(\mathbf{X}, \mu)$ [14, §4]. Thus for f nondihedral, we obtain

$$\frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi) \frac{\mu_f(E(\chi, \cdot))}{\mu_f(1)} \frac{d\chi}{2\pi i}. \tag{2.38}$$

Soundararajan's bound (2.30) for unitary Eisenstein series shows that the right-hand side of (2.38) is

$$\ll_\varepsilon \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(1/2)} \left| \Psi^\wedge(\chi) \frac{C(\chi)^2 \log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)} \right| |d\chi| \ll_\phi \frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)},$$

where in the final step we invoked the rapid decay of $\Psi^\wedge$ (see §2.2.6). Thus we obtain the estimate (2.36) for nondihedral forms over a totally real field.

By combining Holowinsky's theorem 2.3.1 with Soundararajan's (2.33) and (2.36), Holowinsky and Soundararajan obtain, for $\mathbb{F} = \mathbb{Q}$ and $\phi = E(\Psi, \cdot)$, the bound

$$\left| \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} \right| \ll_{\phi,\varepsilon} \min \left(\frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}, \log(k^{\mathbf{1}})^{1/12+\varepsilon} L(\operatorname{ad} f, 1)^{1/4} \right), \tag{2.39}$$

which is $o(1)$ (or even $\ll \log(k^{\mathbf{1}})^{-2/15+\varepsilon}$) by examination (see [25, Proof of Thm 1]). The same estimate follows in the totally real case as soon as one has established theorem 2.3.1.

2.4 The Key Arguments in Our Generalization

We saw in §2.3 that our main result theorem 2.1.1 follows from the generalization of Holowinsky's work asserted by theorem 2.3.1. We now describe the key arguments that reduce our proof of theorem 2.3.1 to several technical results that we shall prove in the remaining sections of this chapter; those results are independent of one another and do not depend upon any work in this section, so there is no circularity in our discussion.

Recall that theorem 2.3.1 claims to bound $\mu_f(\phi)/\mu_f(1) - \mu(\phi)/\mu(1)$, for f a nondihedral holomorphic eigencuspform of weight k and ϕ either a Maass eigencuspform or an incomplete

Eisenstein series, in terms of certain quantities $M_f(k^{\mathbf{1}})$ and $R_f(k^{\mathbf{1}})$ (2.22)–(2.23).

Definition 2.4.1. Fix a nonnegative test function $h \in C_c^\infty(\mathbb{R}_+^*)$ with Mellin transform

$$h^\wedge(s) = \int_0^\infty h(y) y^{-s} \, d^\times y$$

normalized so that $h^\wedge(1) \operatorname{res}_{s=1} E(s,\cdot) = 1$. Recall from §2.2.1 that we have fixed representatives $\mathfrak{z}_j = \operatorname{div} z_j$ for the narrow class group of $\mathbb{F}$; here $j \in \{1, \dots, h(\mathbb{F})\}$ and $z_j \in \mathbb{A}_f^*$. For each unramified idele class character $\chi \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)$ and each $x \geq 2$, define the *shifted sums*

$$S_\chi(x) = \sum_{j=1}^{h(\mathbb{F})} \sum_{\substack{l \in \mathfrak{o}_+^* \backslash \mathfrak{z}_j \\ 0 \neq |l^{\mathbf{1}}| < x^{1+\varepsilon}}} \frac{\lambda_\chi(\mathfrak{z}_j^{-1} l)}{\mathrm{N}(\mathfrak{z}_j^{-1} l)^{1/2}} S_{\chi_\infty}(\mathfrak{z}_j, l, x), \tag{2.40}$$

where

$$S_{\chi_\infty}(\mathfrak{z}, l, x) = \sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^* \\ m := n + l \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^*}} \frac{\lambda_f(\mathfrak{z}^{-1} m)}{\mathrm{N}(\mathfrak{z}^{-1} m)^{1/2}} \frac{\lambda_f(\mathfrak{z}^{-1} n)}{\mathrm{N}(\mathfrak{z}^{-1} n)^{1/2}} \frac{I_{\chi_\infty}(l, n, \mathrm{N}(\mathfrak{z}) x)}{\mathrm{N}(\mathfrak{z})}, \tag{2.41}$$

and (here $m := n + l$ as always)

$$I_{\chi_\infty}(l, n, x) = \frac{(4\pi \mathbf{1})^{k - \mathbf{1}}}{\mathbf{\Gamma}(k - \mathbf{1})} \int_{\mathbb{F}_{\infty+}^*} h(x y^{\mathbf{1}}) \kappa_{\chi,\infty}(ly) \kappa_{f,\infty}(my) \kappa_{f,\infty}(ny) \frac{d^\times y}{y^{\mathbf{1}}}. \tag{2.42}$$

If ϕ is a Maass eigencuspform of eigenvalue $(\frac{1}{4} + r_1^2, \dots, \frac{1}{4} + r_{[\mathbb{F}:\mathbb{Q}]}^2)$ and parity $(\varepsilon_1, \dots, \varepsilon_{[\mathbb{F}:\mathbb{Q}]})$, define analogously $S_\phi(x)$, $S_{\phi_\infty}(\mathfrak{z}, l, x)$ and $I_{\phi_\infty}(l, n, x)$ by replacing $\kappa_{\chi,\infty}$ and λ_χ with $\kappa_{\phi,\infty}$ and λ_ϕ above; note then that $S_{\phi_\infty}(\mathfrak{z}, l, x)$ is the special case of $S_{\chi_\infty}(\mathfrak{z}, l, x)$ obtained by taking χ_∞ to be the (conceivably non-unitary) character $[y \mapsto \prod \operatorname{sgn}(y_j)^{\varepsilon_j} |y_j|^{i r_j}] \in \mathfrak{X}(\mathbb{F}_\infty^*)$ as in (2.5).

Proposition 2.4.2. *Let f be as in the statement of theorem 2.3.1 and let $Y \geq 1$. If ϕ is a Maass eigencuspform, then*

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{c_1(\mathbb{F})}{L(\operatorname{ad} f, 1)} \frac{S_\phi(Y)}{(k - \mathbf{1})^{\mathbf{1}} Y} + O_{\phi,\varepsilon}(Y^{-1/2}).$$

If $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series (recall that f is not dihedral), then

$$\begin{aligned} \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} = \frac{c_1(\mathbb{F})}{L(\operatorname{ad} f, 1)} &\int_{\mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)(0)} \frac{\Psi^\wedge(|.|^{1/2}\chi)}{\xi_\mathbb{F}(|.|\chi^2)\chi(d_\mathbb{F})^{-2}} \frac{S_\chi(Y)}{(k - \mathbf{1})^{\mathbf{1}} Y} \frac{d\chi}{2\pi i} \\ &+ O_{\phi,\varepsilon}\left(\frac{1 + R_f(k^{\mathbf{1}})}{Y^{1/2}}\right). \end{aligned}$$

The constant $c_1(\mathbb{F})$ is as in the formula (2.21).

Proof. See §2.5. The proof is a straightforward and naïve generalization of Holowinsky's arguments in the $\mathbb{F} = \mathbb{Q}$ case. $\square$

Proposition 2.4.2 shows that theorem 2.3.1 follows from sufficiently strong bounds for the shifted sums $S_\phi(Y)$ for ϕ a Maass eigencuspform and $S_\chi(Y)$ for $\chi \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)(0)$ an unramified unitary idele class character.

We bound the sums $S_\phi(Y)$ and $S_\chi(Y)$ by bounding their summands $S_{\chi_\infty}(\mathfrak{z}, l, x)$ for each narrow ideal class representative $\mathfrak{z} = \mathfrak{z}_j$ ($j \in \{1, \dots, \mathbb{F}\}$), each nonzero shift $l \in \mathfrak{z} \cap \mathbb{F}^*$, and each character $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$; recall from Definition 2.4.1 that

$$S_{\phi_\infty}(\mathfrak{z}, l, x) = S_{\chi_\infty}(\mathfrak{z}, l, x) \tag{2.43}$$

for a suitable character $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$. For this reason it suffices to bound $S_{\chi_\infty}(l, n, x)$ when χ_∞ is either unitary or of the form (2.5) for some Maass eigencuspform ϕ, so that in particular each $r_j \in \mathbb{R} \cup i(-\frac{1}{2}, \frac{1}{2})$; we assume henceforth that this is the case.

The sums $S_{\chi_\infty}(\mathfrak{z}, l, x)$ are weighted by an integral $I_{\chi_\infty}(l, n, x)$, which we treat as follows. By the Mellin formula $h(y) = \int_{(c)} h^\wedge(s) y^s \frac{ds}{2\pi i}$ with $h^\wedge(s) = \int_0^\infty h(y) y^{-s} \, d^\times y$ and $c \geq 0$, we may factor $I_{\chi_\infty}(l, n, x)$ as a product of local integrals

$$I_{\chi_\infty}(l, n, x) = \int_{(c)} h^\wedge(s) x^s \left(\prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} J_{ir_j}(l_j, n_j, s) \right) \frac{ds}{2\pi i}, \tag{2.44}$$

where

$$J_{ir_j}(l_j, n_j, s) := \frac{(4\pi)^{k_j - 1}}{\Gamma(k_j - 1)} \int_{\mathbb{R}_+^*} y^{s-1} \kappa_{\chi, \infty_j}(l_j y) \kappa_{f, \infty_j}(m_j y) \kappa_{f, \infty_j}(n_j y) \, d^\times y.$$

The "trivial" bound for J_{ir_j} obtained by applying the inequality $|\kappa_{\chi,\infty_j}(l_j y)| \leq 1$ to the integrand and evaluating the resulting gamma integral is

$$|J_{ir_j}(l_j, n_j, s)| \leq \frac{\Gamma(k_j - 1 + \sigma)}{\Gamma(k_j - 1)} \frac{\sqrt{m_j n_j}}{\left(4\pi \left(\frac{m_j + n_j}{2}\right)\right)^\sigma} \left(\frac{\sqrt{m_j n_j}}{\left(\frac{m_j + n_j}{2}\right)} \right)^{k_j - 1}, \tag{2.45}$$

where $s = \sigma + it$. However, (2.45) would *not suffice* for our purposes, as we shall explain after proving the following refinement.

Lemma 2.4.3. *For $ir_j \in i\mathbb{R} \cup (-\frac{1}{2}, \frac{1}{2})$, $l_j \neq 0$, $n_j > 0$, $m_j = n_j + l_j > 0$, $k_j \geq 2$, and $s = \sigma + it$ with $\sigma \geq -\frac{1}{2}$, we have*

$$|J_{ir_j}(l_j, n_j, s)| \leq \frac{\Gamma(k_j - 1 + \sigma)}{\Gamma(k_j - 1)} \frac{\sqrt{m_j n_j}}{(4\pi \max(m_j, n_j))^\sigma} \left(\frac{\min(m_j, n_j)}{\max(m_j, n_j)} \right)^{\frac{k_j - 1}{2}}. \tag{2.46}$$

Proof. By the integral formula [16, 6.621.3] and the transformation formula [16, 9.131] in

Gradshteyn-Ryzhik, we have explicitly

$$\begin{aligned}J_{ir_j}(l_j,n_j,s) = \pm&\frac{\Gamma(k_j-1+s)}{\Gamma(k_j-1)}\frac{\sqrt{m_jn_j}}{(4\pi\max(m_j,n_j))^s}\left(\frac{\min(m_j,n_j)}{\max(m_j,n_j)}\right)^{\frac{k_j-1}{2}}\\ &\cdot\frac{\Gamma(k_j+s-\frac{1}{2}+ir_j)\Gamma(k_j+s-\frac{1}{2}-ir_j)}{\Gamma(k_j+s-1)\Gamma(k_j+s)}\\ &\cdot {}_2F_1\left(\begin{matrix}\frac{1}{2}-ir_j,\frac{1}{2}+ir_j\\ k_j+s\end{matrix};-\frac{\min(m_j,n_j)}{|m_j-n_j|}\right)\end{aligned} \tag{2.47}$$

where ${}_2F_1$ is the Gauss hypergeometric function and the sign is given by $\prod\operatorname{sgn}(l_j)^{\varepsilon_j}$. By the technical lemmas proved in §2.8, the factors on the second and third lines of (2.47) are each bounded in absolute value by 1, so the claim follows from the basic inequality $|\Gamma(k_j-1+s)|\leq\Gamma(k_j-1+\sigma)$. ☐

Corollary 2.4.4. *Let $\chi_\infty\in\mathfrak{X}(\mathbb{F}_\infty^*)$ be of the form* (2.5) *with each $ir_j\in i\mathbb{R}\cup(-\frac{1}{2},\frac{1}{2})$. Then*

$$I_{\chi_\infty}(l,n,x)\ll_A\sqrt{m^{\mathbf{1}}n^{\mathbf{1}}}\left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)}\right)^{\frac{k-\mathbf{1}}{2}}\min\left(1,\frac{k^{\mathbf{1}}x}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^{A}. \tag{2.48}$$

Proof. Substitute (2.46) into (2.44), taking $c\in\{0,A\}$ and invoking the well known estimate $\Gamma(k_j-1+\sigma)/\Gamma(k_j-1)\ll_\sigma k_j^\sigma$ [72, Ch 7, Misc. Ex 44]. ☐

Remark 1. With more effort (e.g., by studying the asymptotics of the expression (2.47)) one can show that if the components of the weight k increase in such a way that $\min(k_1,\dots,k_{[\mathbb{F}:\mathbb{Q}]})\gg(k^{\mathbf{1}})^{\delta_0}$ for some $\delta_0>0$, then (setting $\mathbf{log}(x)=(\log x_1,\dots,\log x_{[\mathbb{F}:\mathbb{Q}]})$ for $x\in\mathbb{F}_{\infty+}^*\cong(\mathbb{R}_+^*)^{[\mathbb{F}:\mathbb{Q}]}$)

$$\begin{aligned}I_{\chi_\infty}(l,n,x)=\sqrt{m^{\mathbf{1}}n^{\mathbf{1}}}&\left[\kappa_{\chi,\infty}\left(\frac{k-\mathbf{1}}{4\pi}\left|\mathbf{log}\frac{m}{n}\right|\right)h\left(\frac{x\left(\frac{k-\mathbf{1}}{4\pi}\right)^{\mathbf{1}}}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)\right.\\ &\left.+O_{\chi_\infty}\left((k^{\mathbf{1}})^{-\delta_0}\left(\frac{k^{\mathbf{1}}x}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^{1+\varepsilon}\right)\right].\end{aligned}$$

It follows with some work that for ϕ a Maass eigencuspform and $Y\geq 1$, we have

$$\begin{aligned}\frac{\mu_f(\phi)}{\mu_f(1)}=O_\phi(Y^{-1/2})+\frac{c_1(\mathbb{F})}{k^{\mathbf{1}}YL(\operatorname{ad}f,1)}&\sum_{j=1}^{h(\mathbb{F})}\sum_{\substack{l\in\mathfrak{o}_+^*\backslash\mathfrak{z}_j\\ 0\neq|l^{\mathbf{1}}|<Y^{1+\varepsilon}}}\frac{\lambda_\phi(\mathfrak{z}_j^{-1}l)}{\mathrm{N}(\mathfrak{z}_j^{-1}l)^{1/2}}\\ &\cdot\sum_{\substack{n\in\mathfrak{z}\cap\mathbb{F}_{\infty+}^*\\ m:=n+l\in\mathfrak{z}\cap\mathbb{F}_{\infty+}^*}}\lambda_f(\mathfrak{z}^{-1}m)\lambda_f(\mathfrak{z}^{-1}n)\\ &\cdot\kappa_{\phi,\infty}\left(\frac{k-\mathbf{1}}{4\pi}\left|\mathbf{log}\frac{m}{n}\right|\right)\frac{h\left(\frac{Y\,\mathrm{N}(\mathfrak{z})\left(\frac{k-\mathbf{1}}{4\pi}\right)^{\mathbf{1}}}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)}{\mathrm{N}(\mathfrak{z})}.\end{aligned}$$

This refinement is not necessary for our purposes, so we omit the proof; the simpler upper bound given by Corollary 2.4.4 suffices because we do not exploit cancellation in the shifted sums, and has the advantage of being completely uniform in χ_∞.

Corollary 2.4.5. *Let $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$ satisfy the hypotheses of Corollary 2.4.4. Then the shifted sums $S_{\chi_\infty}(\mathfrak{z}, l, Y)$ are bounded up to a multiple depending only upon $\mathfrak{z}$ and A by the quantity*

$$\sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}_{\infty +}^* \\ m := n + l \in \mathfrak{z} \cap \mathbb{F}_{\infty +}^*}} \left| \lambda_f(\mathfrak{z}^{-1} m) \lambda_f(\mathfrak{z}^{-1} n) \right| \left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)} \right)^{\frac{k - \mathbf{1}}{2}} \min\left(1, \frac{k^{\mathbf{1}} Y}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^A. \tag{2.49}$$

Proof. Substitute Corollary 2.4.4 into Definition 2.4.1. □

Remark 2. When $\mathbb{F} = \mathbb{Q}$, Holowinsky applies what amounts to the trivial bound (2.45), which gives something like (2.49) upon replacing

$$\left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)} \right)^{\frac{k - \mathbf{1}}{2}} = \prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} \left(\frac{\min(m_j, n_j)}{\max(m_j, n_j)} \right)^{\frac{k_j - 1}{2}} \quad \text{by} \quad \prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} \left(\frac{\sqrt{m_j n_j}}{\left(\frac{m_j + n_j}{2}\right)} \right)^{\boldsymbol{k} - \mathbf{1}}. \tag{2.50}$$

He then bounds the factor on the RHS of (2.50) by 1. Now, bounding either of the factors in (2.50) is harmless when $\mathbb{F} = \mathbb{Q}$: if f has weight k, then in the sum (2.49) we typically have $m, n \asymp kY$, so for $|l| = O(1)$ both factors in (2.50) are typically $\asymp 1$. On the other hand, when $d = [\mathbb{F} : \mathbb{Q}] > 1$ it is costly to apply such bounds prematurely: the sum (2.49) then has roughly $x \log(x)^{d-1}$ nonnegligible terms with $x = k^{\mathbf{1}} Y$, and this extra logarithmic factor "$\log(x)^{d-1}$" turns out to be unaffordable in the application to mass equidistribution. One can show that the savings obtained by treating nontrivially the factor on the RHS of (2.50) are negligible even for $d > 1$. Thus the success of our method when $\mathbb{F} \neq \mathbb{Q}$ depends crucially on the more careful treatment afforded by Corollary 2.4.4. In fact, the key to our whole argument is that the factor on the LHS of (2.50) is very small if *any component* of $\mathbf{max}(m,n)$ is not too large, as we quantify in Lemma 2.4.7.

Definition 2.4.6. Given parameters $T = (T_1, \ldots, T_d) \in \mathbb{R}_{\geq 1}^{[\mathbb{F}:\mathbb{Q}]}$ and $U \in \mathbb{R}_{\geq 1}$, let

$$\mathcal{R}_{T,U} = \left\{ x \in \mathbb{R}^{[\mathbb{F}:\mathbb{Q}]} : x^{\mathbf{1}} \leq T^{\mathbf{1}}, x \geq T/U \right\}$$

be the subregion of $\mathbb{R}_{>0}^{[\mathbb{F}:\mathbb{Q}]}$ bounded by the hyperbola $\{\prod x_i = \prod T_i\}$ and the hyperplanes $\{x_i = T_i/U\}$. For a multiplicative function $\lambda : I_{\mathbb{F}} \to \mathbb{C}$, an ideal $\mathfrak{z}$ in $\mathbb{F}$ and an element $l \in \mathfrak{z}$, let

$$\Sigma_\lambda(\mathfrak{z}, l, T, U) := \sum_{\substack{n \in \mathfrak{z} \\ m := n + l \in \mathfrak{z} \\ \mathbf{max}(m,n) \in \mathcal{R}_{T,U}}} |\lambda(\mathfrak{z}^{-1} m) \lambda(\mathfrak{z}^{-1} n)|. \tag{2.51}$$

Lemma 2.4.7. *Let $\chi \in \mathfrak{X}(\mathbb{F}^*_{\infty+})$ satisfy the hypotheses of Corollary 2.4.4, let*

$$d = [\mathbb{F} : \mathbb{Q}], \quad T = (T_1, \ldots, T_d) \text{ with } T_i = k_i Y^{1/d}, \quad X = T_1 \ldots T_d = k^{\mathbf{1}} Y,$$

and let $U = \exp(\log(X)^{\varepsilon})$. Suppose that $1 \leq Y \ll \log(X)^{O(1)}$. Then for any ideal $\mathfrak{z}$, any nonzero shift $l \in \mathfrak{z} \cap \mathbb{F}^*$, and any positive integer A, we have

$$S_{\chi_\infty}(l, n, Y) \ll_{\mathfrak{z}, A} X^{-A} + \sum_{r=0}^{\infty} 2^{-rdA} \Sigma_{\lambda_f}(\mathfrak{z}, l, 2^{r+1}T, 2^{r+1}U). \tag{2.52}$$

Proof. We work with the bound asserted by Corollary 2.4.5. Partition those m, n in (2.49) for which $\mathbf{max}(m, n) \geq T/U$ according to the least integer $r \geq 0$ such that $\mathbf{max}(m, n)^{\mathbf{1}} \leq 2^r X$; their contribution is bounded by the second term on the RHS of (2.52). It remains to consider those m, n for which

$$\max(m_i, n_i) \leq T_i/U \tag{2.53}$$

for some index $i \in \{1, \ldots, d\}$. The elementary inequality $1 - x \leq \exp(-x)$ and the tautology $\mathbf{min}(m, n) + |l| = \mathbf{max}(m, n)$ show that

$$\left(\frac{\mathbf{min}(m, n)}{\mathbf{max}(m, n)} \right)^{\frac{\mathbf{k}-\mathbf{1}}{2}} \leq \exp\left(-\sum_{j=1}^{d} \frac{k_j - 1}{2} \frac{|l_j|}{\max(m_j, n_j)} \right),$$

so the assumption (2.53) implies

$$\left(\frac{\mathbf{min}(m, n)}{\mathbf{max}(m, n)} \right)^{\frac{\mathbf{k}-\mathbf{1}}{2}} \leq \exp\left(-\frac{|l_i| U}{3 Y^{1/d}} \right). \tag{2.54}$$

Here we may and shall assume that the shift l is balanced in the sense that $|l_i| \asymp_{\mathfrak{z}} |l_j|$ for all $i, j \in \{1, \ldots, \mathbb{F}\}$ since $S_{\chi_\infty}(\eta l, n, Y) = S_{\chi_\infty}(l, n, Y)$ for any totally positive unit $\eta \in \mathfrak{o}^*_+$; in particular, we may assume that there exists a positive number c, depending only upon the fixed number field $\mathbb{F}$ and the fixed set of representatives $\{\mathfrak{z}_1, \ldots, \mathfrak{z}_{h(\mathbb{F})}\}$ for the narrow class group, such that $|l_i| \geq c$ for each i. Since $Y \ll \log(X)^{O(1)}$ by assumption, our choice $U = \exp(\log(X)^{\varepsilon})$ is (more than) large enough that for each positive real A the inequality

$$\frac{cU}{3Y^{1/d}} \geq A \log(X)$$

holds eventually (i.e., for $\max(k_1, \ldots, k_d) \gg 1$), so by (2.54) we obtain

$$\left(\frac{\mathbf{min}(m, n)}{\mathbf{max}(m, n)} \right)^{\frac{\mathbf{k}-\mathbf{1}}{2}} \ll_A X^{-A}. \tag{2.55}$$

By the trivial "Hecke" bound $\lambda_f(\mathfrak{a}) \ll \mathrm{N}(\mathfrak{a})^{1/2+\varepsilon}$, the contribution to (2.49) of n satisfying (2.53) is

$$\begin{aligned} &\ll X^{-A'} \sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+} \\ m := n+l \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+}}} \left|\lambda_f(\mathfrak{z}^{-1}m)\lambda_f(\mathfrak{z}^{-1}n)\right| \min\left(1, \frac{X}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^A \\ &\ll X^{-A'} \sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+} \\ m := n+l \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+}}} (m^{\mathbf{1}} n^{\mathbf{1}})^{1/2+\varepsilon} \min\left(1, \frac{X}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^A \end{aligned} \tag{2.56}$$

for any $A, A' > 0$. Since $|l|_i \geq c$, the number of $n \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+}$ for which $n + l \in \mathfrak{z} \cap \mathbb{F}^*_{\infty+}$ and $\mathbf{max}(m,n)^{\mathbf{1}} \leq 2^r X$ $(r \geq 0)$ is $\ll (2^r X)^d$. Choosing $A = 1 + 2\varepsilon + d + 1$, summing dyadically, and taking A' to be sufficiently large, we see that (2.56) is $\ll_{A''} X^{-A''}$ for any positive constant A'', as desired. ☐

The volume of $\mathcal{R}_{T,U}$ is approximately $X \log(U)^{d-1} = X \log(X)^{(d-1)\varepsilon}$. Since the number of nonnegligible terms appearing in $S_{\chi_\infty}(l, n, Y)$ is approximately $X \log(X)^{d-1}$, we see that Lemma 2.4.7 allows us to discard the vast majority of those terms. We treat the remaining $\approx X \log(X)^{\varepsilon'}$ terms by the following generalization of Holowinsky's bound for shifted sums of multiplicative functions [24, Thm 2].

Theorem 2.4.8. *Let $T \in \mathbb{R}_{\geq 1}^{[\mathbb{F}:\mathbb{Q}]}$, $U \in \mathbb{R}_{\geq 1}$, $\mathfrak{z}$, l and $\lambda : I_{\mathbb{F}} \to \mathbb{C}$ be as in Definition 2.4.6. Suppose that $l \neq 0$ and that $|\lambda(\mathfrak{a})| \leq \tau(\mathfrak{a})$ for all integral ideals $\mathfrak{a}$. Set $X = T^{\mathbf{1}}$ and $d = [\mathbb{F} : \mathbb{Q}]$. Then*

$$\Sigma_\lambda(\mathfrak{z}, l, T, U) \ll_{\mathfrak{z},\varepsilon} \frac{\log(eU)^{d-1} X}{\log(eX)^{2-\varepsilon}} \prod_{\mathrm{N}(\mathfrak{p}) \leq X} \left(1 + \frac{2|\lambda(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right). \tag{2.57}$$

Here the product is taken over prime ideals of norm at most X.

Proof. See §2.6. ☐

Remark 3. Holowinsky [24, Thm 2] established a slightly weaker form of the case $d = 1$ of theorem 2.4.8 by an application of the large sieve; in his inequality (2.2) an additional factor of $\tau(l)$ appears on the RHS. We prove theorem 2.4.8 by adapting his approach, with the only difficulty being that the regions $\mathcal{R}_{T,U}$ are shaped quite differently when $d > 1$.

If one is willing to sacrifice uniformity in the shift l, then alternate proofs of the corresponding weakening of Holowinsky's [24, Thm 2] and (probably) our theorem 2.4.8 can be obtained by the general estimates due to Nair [45] and Nair-Tenenbaum [46] for sums $\sum_n \lambda(|P(n)|)$ with P a (primitive, possibly multivariate) polynomial (for example, $P(n) = n(n+l)$) and n traversing a box; note that in all of the bounds asserted by Nair and Nair-Tenenbaum, the implied constants depend in an unspecified manner upon the discriminant and degree of P. This seems insufficient in the application to QUE where the shift l must vary (particularly when ϕ is an incomplete Eisenstein series, see [64]).

We refer to [47, Rmk 3.11] for a further discussion of variations on the $d = 1$ case of theorem 2.4.8 that may be derived from other works and particularly their applicability to QUE in the level aspect.

Proof of theorem 2.3.1. Let $Y \geq 1$ be a parameter (to be chosen at the end of the proof) that satisfies $Y \ll \log(k^{\mathbf{1}})^{O(1)}$. Preserve the hypotheses and notation $d = [\mathbb{F} : \mathbb{Q}]$, $T = Y^{1/d}k$, $X = T^{\mathbf{1}} = k^{\mathbf{1}}Y$ and $U = \exp(\log(X)^{\varepsilon})$ from above. Lemma 2.4.7 and theorem 2.4.8 show that

$$S_{\chi_\infty}(l,n,Y) \ll_{A,\varepsilon} X^{-A} + \sum_{r=0}^{\infty} 2^{-rdA} \frac{\log(2^r eU)^{d-1} 2^{rd} X}{\log(2^{rd}X)^{2-\varepsilon}} \prod_{\mathrm{N}(\mathfrak{p}) \leq 2^r X} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right). \tag{2.58}$$

Taking $A = 2$ and using that

$$\sum_{r=0}^{\infty} 2^{rd-rdA} \log(2^r eU)^{d-1} \prod_{X < \mathrm{N}(\mathfrak{p}) \leq 2^r X} \left(1 + \frac{4}{\mathrm{N}(\mathfrak{p})}\right) \ll_{\varepsilon} \log(X)^{(d-1)\varepsilon}$$

gives

$$S_{\chi_\infty}(l,n,Y) \ll_{\varepsilon} \frac{X}{\log(X)^{2-\varepsilon'}} \prod_{\mathrm{N}(\mathfrak{p}) \leq X} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right),$$

where $\varepsilon' = d\varepsilon$. Thus

$$S_{\phi}(Y) \ll_{\phi,\varepsilon} \frac{k^{\mathbf{1}} Y^{3/2+\varepsilon}}{\log(k^{\mathbf{1}})^{2-\varepsilon'}} \prod_{\mathrm{N}(\mathfrak{p}) \leq k^{\mathbf{1}}} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right), \tag{2.59}$$

since the sum over l in Definition 2.4.1 introduces the additional factor

$$\sum_{\substack{0 \neq \mathfrak{a} \subset \mathfrak{o} \\ \mathrm{N}(\mathfrak{a}) < Y^{1+\varepsilon}}} \frac{|\lambda_\phi(\mathfrak{a})|}{\mathrm{N}(\mathfrak{a})^{1/2}} \leq \left(\sum_{\substack{0 \neq \mathfrak{a} \subset \mathfrak{o} \\ \mathrm{N}(\mathfrak{a}) < Y^{1+\varepsilon}}} |\lambda_\phi(\mathfrak{a})|^2 \sum_{\substack{0 \neq \mathfrak{b} \subset \mathfrak{o} \\ \mathrm{N}(\mathfrak{b}) < Y^{1+\varepsilon}}} \frac{1}{\mathrm{N}(\mathfrak{b})} \right)^{1/2} \ll_{\phi} Y^{1/2+\varepsilon}$$

by the Cauchy-Schwarz inequality and the Rankin-Selberg bound (2.13); similarly, using that $|\lambda_\chi(\mathfrak{a})| \leq \tau(\mathfrak{a})$ for a *unitary* character $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)$, we find that

$$S_{\chi}(Y) \ll_{\varepsilon} \frac{k^{\mathbf{1}} Y^{3/2+\varepsilon}}{\log(k^{\mathbf{1}})^{2-\varepsilon'}} \prod_{\mathrm{N}(\mathfrak{p}) \leq k^{\mathbf{1}}} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right), \tag{2.60}$$

where we emphasize that the implied constant does not depend upon χ. By Proposition 2.4.2 and the definitions (2.22)–(2.23) of $M_f(x)$ and $R_f(x)$, we deduce for ϕ a Maass eigencuspform that

$$\frac{\mu_f(\phi)}{\mu_f(1)} \ll_{\phi,\varepsilon} Y^{1/2+\varepsilon} \log(k^{\mathbf{1}})^{\varepsilon'} M_f(k^{\mathbf{1}}) \tag{2.61}$$

and for $\phi = E(\Psi, \cdot)$ an incomplete Eisenstein series that

$$\begin{aligned}\frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} \ll_{\phi,\varepsilon} Y^{1/2+\varepsilon} \log(k^{\mathbf{1}})^{\varepsilon'} M_f(k^{\mathbf{1}}) \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)} \left| \frac{\Psi^\wedge(|.|^{1/2}\chi)}{\xi_{\mathbb{F}}(|.|^1\chi^2)} \right| |d\chi| \\ + \frac{1 + R_f(k^{\mathbf{1}})}{Y^{1/2}}.\end{aligned} \tag{2.62}$$

The integral in (2.62) converges by the rapid decay of $\Psi^\wedge$ (see §2.2.6). Choosing (as Holowinsky does) $Y = \max(1, M_f(k^{\mathbf{1}})^{-1}) \ll \log(k^{\mathbf{1}})^{O(1)}$ in (2.61) and (2.62), we conclude the proof of theorem 2.3.1. □

2.5 Reduction to Shifted Sums Weighted by an Integral

In this section we establish Proposition 2.4.2, which reduces our study of $\mu_f(\phi)$ to that of the shifted sums $S_\phi(Y)$ and $S_\chi(Y)$; here and throughout this section $Y \geq 1$ is a (small) parameter, f is a nondihedral holomorphic eigencuspform of weight $k = (k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$, ϕ is a Maass eigencuspform or incomplete Eisenstein series, and $h \in C_c^\infty(\mathbb{R}_+^*)$ is a fixed test function with Mellin transform $h^\wedge(s) = \int_0^\infty h(y) y^{-s} \, d^\times y$ normalized as in Definition 2.4.1 so that

$$h^\wedge(1) \operatorname{res}_{s=1} E(s, \cdot) = 1. \tag{2.63}$$

Let h_Y be the function $y \mapsto h(Yy)$ and let

$$E(h_Y, \cdot) : G(\mathbb{A}) \ni g \mapsto \sum_{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F})} h_Y(|y(\gamma g)|)$$

be the incomplete Eisenstein series attached by the recipe of §2.2.8.4 to the test function $h_Y \circ |.| \in C_c^\infty(C_{\mathbb{F}}/C_{\mathbb{F}}^1) \hookrightarrow C_c^\infty(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$.

Lemma 2.5.1. *We have the approximate formula*

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{\mu_f(E(h_Y, \cdot)\phi)}{Y \mu_f(1)} + O_\phi(Y^{-1/2}).$$

Proof. The starting point is the consequence

$$\mu_f(E(h_Y, \cdot)\phi) = Y \mu_f(\phi) + \int_{(1/2)} h_Y^\wedge(s) \mu_f(E(s, \cdot)\phi) \, \frac{ds}{2\pi i}, \tag{2.64}$$

of Mellin inversion, Cauchy's theorem and our normalization (2.63). We need a crude bound of the form

$$E(s,g)\phi(g) \ll_\phi |s|^{2[\mathbb{F}:\mathbb{Q}]+\varepsilon} \quad \text{for } \mathrm{Re}(s) = \tfrac{1}{2}, g \in G(\mathbb{A}), \tag{2.65}$$

where the precise exponent is not important. To establish this, recall first that if $c > 0$ is chosen small enough, then the Siegel set $\mathfrak{S}$ consisting of those $g = n(x)a(y)kz \in G(\mathbb{A})$ for which $|y| \geq c$ satisfies $G(\mathbb{A}) = G(\mathbb{F})\mathfrak{S}$. Since $E(s,\cdot)\phi$ is $Z(\mathbb{A})$-invariant and right K-invariant, it suffices to establish (2.65) for $g = n(x)a(y \times z_j^{-1})$ where $x \in \mathbb{A}$, $y \in \mathbb{F}_{\infty+}^*$ with $y^{\mathbf{1}} \geq c$ and $j \in \{1,\ldots,h(\mathbb{F})\}$. For $s = \frac{1}{2} + it$ the Fourier expansion of $E(s,\cdot)$, given in §2.2.8.3, shows that

$$|E(s, n(x)a(y \times z_j^{-1})))| \ll (y^{\mathbf{1}})^{1/2} + \sum_{n \in \mathbb{F}^* \cap \mathfrak{z}_j} \left| \frac{\kappa_{it,\infty}(ny)}{\xi_{\mathbb{F}}(1+2it)} \frac{\lambda_{it}(\mathfrak{z}_j^{-1}n)}{\mathrm{N}(\mathfrak{z}_j^{-1}n)^{1/2}} \right|, \tag{2.66}$$

where for simplicity we write $\kappa_{it,\infty} := \kappa_{|.|^{it},\infty}$ and $\lambda_{it} := \lambda_{|.|^{it}}$. The straightforward analysis of [67, §3.6] applied to $\zeta_{\mathbb{F}}$ in place of $\zeta_{\mathbb{Q}}$ shows that[4]

$$\xi_{\mathbb{F}}(1+2it)^{-1} \ll \frac{(1+|t|)^\varepsilon}{\Gamma_{\mathbb{R}}(1+2it)^{[\mathbb{F}:\mathbb{Q}]}},$$

and it is noted in [24, page 6] that the integral formula for K_{it} implies

$$\frac{K_{it}(y)}{\Gamma_{\mathbb{R}}(1+2it)} \ll \left(\frac{1+|t|}{y}\right)^A \left(1 + \frac{1+|t|}{y}\right)^\varepsilon \quad \text{for any } A \in \mathbb{Z}_{\geq 0},\ \varepsilon > 0,$$

thus (writing $d = [\mathbb{F}:\mathbb{Q}]$, $\varepsilon' = (d+1)\varepsilon$, and using that $|n^{\mathbf{1}}|y^{\mathbf{1}} \gg 1$)

$$\left| \frac{\kappa_{it,\infty}(ny)}{\xi_{\mathbb{F}}(1+2it)} \frac{\lambda_{it}(\mathfrak{z}_j^{-1}n)}{\mathrm{N}(\mathfrak{z}_j^{-1}n)^{1/2}} \right| \ll (y^{\mathbf{1}})^{1/2}(1+|t|)^{2d+\varepsilon'} \frac{|n^{\mathbf{1}}|^\varepsilon}{(\mathbf{max}(\mathbf{1}, |n|y)^{\mathbf{1}})^A}.$$

Take $A = 2$. We have

$$\sum_{n \in \mathbb{F}^* \cap \mathfrak{z}_j} \frac{|n^{\mathbf{1}}|^\varepsilon}{(\mathbf{max}(\mathbf{1}, |n|y)^{\mathbf{1}})^2} \ll (y^{\mathbf{1}})^{-2} \tag{2.67}$$

because the LHS of (2.67) is invariant under multiplying y by an element of $\mathfrak{o}_+^*$, so we may assume that y is balanced ($y_i \asymp y_j$ for all i,j) with each component bounded uniformly from below, in which case (2.67) may be compared with a convergent integral. Thus $|E(s, n(x)a(y \times z_j^{-1}))| \ll (y^{\mathbf{1}})^{1/2} + |s|^{2d+\varepsilon'}(y^{\mathbf{1}})^{-3/2}$. Since ϕ satisfies[5] $\phi(n(x)a(y \times z_j^{-1})) \ll_\phi (y^{\mathbf{1}})^{-A}$, we obtain the crude

[4]We believe that the stronger bound with $(1+|t|)^\varepsilon$ replaced by $\log(1+|t|)$ holds, but could not quickly locate a reference.

[5]For a Maass eigencuspform, this is well known [34, Prop 10.7]; an incomplete Eisenstein series vanishes off a compact subset of $\mathbf{X}$.

bound (2.65).

By the rapid decay of $h^\wedge$ and the identity $h_Y^\wedge(s) = Y^s h^\wedge(s)$, we deduce from (2.65) that the error term in (2.64) satisfies

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

The lemma follows upon dividing through by $Y\mu_f(1)$. $\square$

Fix now a nice fundamental domain $[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]$ for the quotient $\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*$ with the property that $y \in [\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]$ implies $y_i \asymp y_j$ for all $i,j \in \{1,\dots,[\mathbb{F}:\mathbb{Q}]\}$. Write the Fourier expansions of ϕ and f in the form

$$\phi = \sum_{l\in\mathbb{F}} \phi_l, \quad f = \sum_{n\in\mathbb{F}^*} f_n, \tag{2.68}$$

where $\phi_l : G(\mathbb{A}) \to \mathbb{C}$ satisfies $\phi_l(n(x)g) = e_{\mathbb{F}}(lx)\phi_l(g)$ for all $x \in \mathbb{A}$ and f_n satisfies the analogous condition.

Lemma 2.5.2. *We have $\mu_f(E(h_Y,\cdot)\phi) = \mathcal{S}_0 + \mathcal{S}_1 + \mathcal{S}_2$, where*

$$\mathcal{S}_0 = \sum_{j=1}^{h(\mathbb{F})} \int_{y\in[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]} \frac{h_Y(y^{\mathbf{1}}\,\mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} \int_{x\in\mathbb{F}\backslash\mathbb{A}} (\phi_0|f|^2)(n(x)a(y\times z_j^{-1}))\,dx\,\frac{d^\times y}{y^{\mathbf{1}}}; \tag{2.69}$$

for ϕ a Maass eigencuspform,

$$\mathcal{S}_1 = \frac{\mathbf{\Gamma}(k-\mathbf{1})}{(4\pi\mathbf{1})^{k-\mathbf{1}}} S_\phi(Y);$$

for $\phi = E(\Psi,\cdot)$ an incomplete Eisenstein series,

$$\mathcal{S}_1 = \frac{\mathbf{\Gamma}(k-\mathbf{1})}{(4\pi\mathbf{1})^{k-\mathbf{1}}} \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)} \frac{\Psi^\wedge(|.|^{1/2}\chi)}{\xi_{\mathbb{F}}(|.|\chi^2)\chi(d_{\mathbb{F}})^{-2}} S_\chi(Y)\,\frac{d\chi}{2\pi i};$$

and

$$|\mathcal{S}_2| \leq \mu_f(E(h_Y,\cdot)) \sum_{j=1}^{h(\mathbb{F})} \sup_{\substack{y\in[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*] \\ h_Y(y^{\mathbf{1}}N(\mathfrak{z}_j))\neq 0}} \sum_{\substack{l\in\mathfrak{z}_j \\ |l^{\mathbf{1}}|\geq Y^{1+\varepsilon}}} |\phi_l(a(y\times z_j^{-1}))|. \tag{2.70}$$

The shifted sums $S_\phi(Y)$ and $S_\chi(Y)$ are as in Definition 2.4.1.

Proof. By the formula (2.4) for integration over $Z(\mathbb{A})B(\mathbb{F})\backslash G(\mathbb{A})$, we see that

$$\begin{aligned}&\mu_f(E(h_Y,\cdot)\phi) \\ &\quad= \sum_{j=1}^{h(\mathbb{F})} \int_{y\in\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*} \frac{h_Y(y^{\mathbf{1}}\,\mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} \int_{x\in\mathbb{F}\backslash\mathbb{A}} (\phi|f|^2)(n(x)a(y\times z_j^{-1}))\,dx\,\frac{d^\times y}{y^{\mathbf{1}}}.\end{aligned} \tag{2.71}$$

We now integrate in y over the fundamental domain $[\mathbb{F}^*_{\infty+}/\mathfrak{o}^*_+]$ and substitute for ϕ its Fourier series $\sum \phi_l$. Note that $\phi_l(n(x)a(y \times z_j^{-1})) = 0$ unless $l \in \mathfrak{z}_j$. The contribution to (2.71) of the constant term ϕ_0 is precisely $\mathcal{S}_0$. Let $\mathcal{S}_2$ denote the contribution of those ϕ_l for which $|l^{\mathbf{1}}| \geq Y^{1+\varepsilon}$, so that the bound (2.70) follows from the formula for $\mu_f(E(h_Y, \cdot))$ given by (2.71) with $\phi = 1$. Let $\mathcal{S}_1$ denote the remaining contribution of those $l \in \mathfrak{z}_j$ for which $0 \neq |l^{\mathbf{1}}| < Y^{1+\varepsilon}$. Substituting the Fourier series $f = \sum f_n$ (in which $f_n(y \times z_j^{-1}) = 0$ unless $n \in \mathfrak{z}_j \cap \mathbb{F}^*_{\infty+}$) and integrating in x, we obtain

$$\mathcal{S}_1 = \sum_{j=1}^{h(\mathbb{F})} \sum\sum_{\substack{(l,n) \in (\mathbb{F}^* \cap \mathfrak{z}_j)^2 \\ l^{\mathbf{1}} < Y^{1+\varepsilon} \\ n \in \mathbb{F}^*_{\infty+} \\ m := n + l \in \mathbb{F}^*_{\infty+}}} \int_{y \in [\mathbb{F}^*_{\infty+}/\mathfrak{o}^*_+]} \frac{h_Y(y^{\mathbf{1}} \operatorname{N}(\mathfrak{z}_j))}{\operatorname{N}(\mathfrak{z}_j)} (\phi_l \overline{f_m} f_n)(a(y \times z_j^{-1})) \frac{d^\times y}{y^{\mathbf{1}}}. \tag{2.72}$$

If $\eta \in \mathfrak{o}^*_+$, then $(\phi_{\eta l} \overline{f_{\eta m}} f_{\eta n})(a(y \times z_j^{-1})) = (\phi_l \overline{f_m} f_n)(a(\eta y \times z_j^{-1}))$ (see §2.2.7), so we may break the sum into orbits for (l, n) under the diagonal action of $\mathfrak{o}^*_+$ and unfold the integral over y to all of $\mathbb{F}^*_{\infty+}$:

$$\mathcal{S}_1 = \sum_{j=1}^{h(\mathbb{F})} \sum\sum_{\substack{(l,n) \in \mathfrak{o}^*_+ \backslash (\mathbb{F}^* \cap \mathfrak{z}_j)^2 \\ l^{\mathbf{1}} < Y^{1+\varepsilon} \\ n \in \mathbb{F}^*_{\infty+} \\ m := n + l \in \mathbb{F}^*_{\infty+}}} \int_{y \in \mathbb{F}^*_{\infty+}} \frac{h_Y(y^{\mathbf{1}} \operatorname{N}(\mathfrak{z}_j))}{\operatorname{N}(\mathfrak{z}_j)} (\phi_l \overline{f_m} f_n)(a(y \times z_j^{-1})) \frac{d^\times y}{y^{\mathbf{1}}}. \tag{2.73}$$

Take as representatives for $\mathfrak{o}^*_+ \backslash (\mathbb{F}^* \cap \mathfrak{z}_j)^2$ the pairs (l, n) with l traversing any set of representatives for $\mathfrak{o}^*_+ \backslash (\mathbb{F}^* \cap \mathfrak{z}_j)$ and n traversing the set $\mathbb{F}^* \cap \mathfrak{z}_j$. Recalling the formulas for f_n and ϕ_l given in §2.2.8.1, §2.2.8.2 and §2.2.8.4 and the definitions of $S_\phi(Y)$ and $S_\chi(Y)$, we obtain the claimed expressions for $\mathcal{S}_1$. $\square$

Lemma 2.5.3. *We have*

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

where $\delta_\phi = 0$ *or* 1 *according as* ϕ *is a Maass eigencuspform and or an incomplete Eisenstein series.*

Proof. If ϕ is cuspidal, then $\mathcal{S}_0 = \mu(\phi) = 0$, so there is nothing to show. Suppose that $\phi = E(\Psi, \cdot)$. If $y^{\mathbf{1}} \asymp Y^{-1}$, then it follows from (2.19) that

$$\phi_0(y \times z_j^{-1}) = \frac{\mu(\phi)}{\mu(1)} + \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0) \chi_0(y \times z_j^{-1}) + O_\phi(Y^{-1/2}). \tag{2.74}$$

We have

$$\begin{aligned} &\sum_{j=1}^{h(\mathbb{F})} \int_{y \in [\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]} \frac{h_Y(y^{\mathbf{1}} \mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} \int_{x \in \mathbb{F} \backslash \mathbb{A}} |f|^2(n(x)a(y \times z_j^{-1}))\, dx\, \frac{d^\times y}{y^{\mathbf{1}}} \\ &\quad = \mu_f(E(h_Y, \cdot)) = \int_{(2)} h_Y^\wedge(s) \mu_f(E(s, \cdot))\, \frac{ds}{2\pi i}, \end{aligned} \tag{2.75}$$

and similarly for $1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]$,

$$\begin{aligned} &\sum_{j=1}^{h(\mathbb{F})} \int_{y \in [\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]} \frac{h_Y(y^{\mathbf{1}} \mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} \chi_0(y \times z_j^{-1}) \int_{x \in \mathbb{F} \backslash \mathbb{A}} |f|^2(n(x)a(y \times z_j^{-1}))\, dx\, \frac{d^\times y}{y^{\mathbf{1}}} \\ &\quad = \int_{(2)} h_Y^\wedge(s) \mu_f(E(|.|^s \chi_0, \cdot))\, \frac{ds}{2\pi i}. \end{aligned} \tag{2.76}$$

Substituting (2.74) into (2.69) and applying (2.75) and (2.76), we obtain

$$\begin{aligned} \mathcal{S}_0 = &\left(\frac{\mu(\phi)}{\mu(1)} + O_\phi(Y^{-1/2}) \right) \int_{(2)} h_Y^\wedge(s) \mu_f(E(s, \cdot))\, \frac{ds}{2\pi i} \\ &+ \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0) \int_{(2)} h_Y^\wedge(s) \mu_f(E(|.|^s \chi_0, \cdot))\, \frac{ds}{2\pi i}. \end{aligned} \tag{2.77}$$

Shift the contours in (2.77) to the line $\mathrm{Re}(s) = \frac{1}{2}$; for $\chi_0 \neq 1$ we do not pick up a pole of $\mu_f(E(|.|^s \chi_0, \cdot))$ because f is nondihedral. Thus

$$\begin{aligned} \mathcal{S}_0 = &\, Y \mu_f(1) \left(\frac{\mu(\phi)}{\mu(1)} + O_\phi(Y^{-1/2}) \right) \\ &+ O_\phi \left(\sum_{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} \int_{(1/2)} |h_Y^\wedge(s) \mu_f(E(\chi_0 |.|^s, \cdot))|\, |ds| \right). \end{aligned} \tag{2.78}$$

To simplify the error term, we apply the formula

$$\begin{aligned} &\frac{\mu_f(E(\chi_0|.|^s, \cdot))}{\mu_f(1)} \\ &\quad = c_1(\mathbb{F}) \int_{(1/2)} h^\wedge(s) \left(\frac{Y}{4\pi^{[\mathbb{F}:\mathbb{Q}]}} \right)^s \frac{\mathbf{\Gamma}(k + (s-1)\mathbf{1})}{\mathbf{\Gamma}(k)} \frac{\zeta_{\mathbb{F}}(\chi_0|.|^s)}{\zeta_{\mathbb{F}}(2s)} \frac{L(\mathrm{ad}\, f, \chi_0|.|^s)}{L(\mathrm{ad}\, f, 1)} \frac{ds}{2\pi i} \end{aligned} \tag{2.79}$$

which follows from the unfolding method and analytic continuation as in §2.2.9. By the standard estimates $|\Gamma(k_j - \frac{1}{2} + it)| \leq \Gamma(k_j - \frac{1}{2}) \ll k_j^{-1/2} \Gamma(k_j)$, $\zeta_{\mathbb{F}}(\chi_0|.|^s) \ll |s|^{[\mathbb{F}:\mathbb{Q}]/4}$ and $|\zeta_{\mathbb{F}}(2s)| \gg |s|^{-\varepsilon}$ for $\mathrm{Re}(s) = \frac{1}{2}$ (see also Soundararajan's arguments [66, p7] when $\mathbb{F} = \mathbb{Q}$), we deduce that the error term in (2.78) satisfies

$$\sum_{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} \int_{(1/2)} |h_Y^\wedge(s) \mu_f(E(\chi_0|.|^s, \cdot))|\, |ds| \ll Y^{1/2} \mu_f(1) R_f(k^{\mathbf{1}}), \tag{2.80}$$

with R_f given by (2.23). The lemma follows upon dividing through by $Y\mu_f(1)$. $\square$

Lemma 2.5.4. *We have*

$$\frac{|\mathcal{S}_2|}{Y\mu_f(1)} \ll Y^{-10}.$$

Proof. Set $d = [\mathbb{F} : \mathbb{Q}]$, and note that each l arising in the sum (2.70) satisfies

$$2^r(Y^{1+\varepsilon})^{1/d} \leq \max(|l_1|, \dots, |l_d|) < 2^{r+1}(Y^{1+\varepsilon})^{1/d} \tag{2.81}$$

for some nonnegative integer r. More generally, there are $\ll 2^{rd}Y^{1+\varepsilon}$ elements $l \in \mathfrak{z}_j$ for which (2.81) holds. For each $y \in [\mathbb{F}^*_{\infty+}/\mathfrak{o}^*_+]$ such that $h_Y(y^{\mathbf{1}}N(\mathfrak{z}_j)) \neq 0$, we have $y^{\mathbf{1}} \asymp Y^{-1}$ and $y_i \asymp y_j$ for $i, j \in \{1, \dots, [\mathbb{F} : \mathbb{Q}]\}$, thus

$$y_i \asymp Y^{-1/d} \quad \text{for each } i. \tag{2.82}$$

Suppose that ϕ is a Maass eigencuspform, so that

$$\phi_l(a(y \times z_j^{-1})) = \kappa_{\phi,\infty}(ly)\frac{\lambda_\phi(lz_j^{-1})}{\mathrm{N}(lz_j^{-1})^{1/2}}.$$

We have $\lambda_\phi(\mathfrak{a}) \leq \tau(\mathfrak{a})\,\mathrm{N}(\mathfrak{a})^{1/2} \ll \mathrm{N}(\mathfrak{a})^{1/2+\varepsilon}$ and $\kappa_{\phi,\infty}(ly) = \prod_{i=1}^d \kappa_{\phi,\infty_i}(l_iy_i)$ with

$$\kappa_{\phi,\infty_i}(l_iy_i) = \pm 2(|l_i|y_i)^{1/2}K_{ir_i}(2\pi|l_i|y_i),$$

where $|\kappa_{\phi,\infty_i}(l_iy_i)| \leq 1$ and

$$K_{ir}(x) \ll \left(\frac{1+|r|}{x}\right)^{A'} \quad \text{uniformly for } r \in \mathbb{R} \cup i(-\tfrac{1}{2}, \tfrac{1}{2}) \text{ and } x \geq \delta > 0. \tag{2.83}$$

Thus if $l \in \mathfrak{z}_j$ and $y \in \mathbb{F}^*_{\infty+}$ satisfy (2.81)–(2.82), we obtain

$$|\phi_l(a(y \times z_j^{-1}))| \ll (1 + |r|^{\mathbf{1}})^{O(1)}(2^rY^{\varepsilon/d})^{-A} \tag{2.84}$$

for any positive A. The dependence of the bound (2.84) on ϕ is polynomial in the archimedean parameters r_i, so (2.84) extends to the case that $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series by the integral formula (2.20) for its Fourier coefficients and the rapid decay of the test function $\Psi^\wedge$.

Taking A sufficiently large in (2.84) and summing over $l \in \mathfrak{z}_j$ that satisfy the condition (2.81) for some $r \in \mathbb{Z}_{\geq 0}$, we deduce

$$|\mathcal{S}_2| \ll Y^{-12}\mu_f(E(h_Y, \cdot)). \tag{2.85}$$

The function h is bounded, so

$$E(h_Y, g) = \sum_{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F})} h(Y|y(\gamma g)|) \ll \#\{\gamma \in B(\mathbb{F})\backslash G(\mathbb{F}) : |y(\gamma g)| \asymp Y^{-1}\}. \tag{2.86}$$

By [68, Lem 8.7], the cardinality on the RHS of (2.86) is $\ll Y^{1+\varepsilon}$, uniformly in g. Thus $E(h_Y, \cdot) \ll Y^{1+\varepsilon}$ and $\mu_f(E(h_Y, \cdot)) \ll Y^{1+\varepsilon}\mu_f(1)$, so (2.85) gives $|\mathcal{S}_2| \ll Y^{-10}\mu_f(1)$. □

Proof of Proposition 2.4.2. Follows immediately from the sequence of lemmas proved in this section together with the consequence

$$\frac{1}{Y\mu_f(1)} \frac{\mathbf{\Gamma}(k - \mathbf{1})}{(4\pi\mathbf{1})^{k-\mathbf{1}}} = \frac{c_1(\mathbb{F})}{L(\mathrm{ad}\, f, 1)} \frac{1}{(k - \mathbf{1})^{\mathbf{1}} Y}$$

of the formula (2.21). □

Remark 4. Let us point out the essential difference between our method and that of Marshall [43]. Recall that starting from Lemma 2.5.1, we have integrated $\phi|f|^2$ against the incomplete Eisenstein series $E(h, \cdot)$ attached to a test function $h \in C_c^\infty(C_\mathbb{F}/C_\mathbb{F}^1) = C_c^\infty(\mathbb{R}_+^*)$. Marshall instead integrates against what he calls a "unipotent Eisenstein series," which (reinterpreted adelically) amounts to the incomplete Eisenstein series $E(H, \cdot)$ attached to the test function $H \in C_c^\infty(C_\mathbb{F}/\hat{\mathfrak{o}}^*)$ given by $H(y) = \sum_{\alpha \in \mathbb{F}^*} h(\alpha y)$ for some pure tensor $h = \prod h_v \in C_c^\infty(\mathbb{A}^*/\hat{\mathfrak{o}}^*)$. Suppose that ϕ is cuspidal; the case that $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series proceeds similarly after separating out the constant term and appealing to the formula (2.20). Then

$$\begin{aligned}
\mu_f(E(H, \cdot)\phi) &= \int_{Z(\mathbb{A})B(\mathbb{F})\backslash G(\mathbb{A})} H\phi|f|^2 \\
&= \int_{y \in \mathbb{F}^* \backslash \mathbb{A}^*} \left(\sum_{\alpha \in \mathbb{F}^*} h(\alpha y) \right) \int_{x \in \mathbb{F}\backslash\mathbb{A}} (\phi|f|^2)(n(x)a(y))\, dx \, \frac{d^\times y}{|y|} \\
&= \int_{y \in \mathbb{A}^*} h(y) \int_{x \in \mathbb{F}\backslash\mathbb{A}} (\phi|f|^2)(n(x)a(y))\, dx \, \frac{d^\times y}{|y|} \\
&= \sum\sum_{\substack{(l,n) \in \mathbb{F}^* \times \mathbb{F}^* \\ m := n + l \in \mathbb{F}^*}} \int_{y \in \mathbb{A}^*} h(y)\kappa_\phi(ly)\kappa_f(my)\kappa_f(ny) \, \frac{d^\times y}{|y|}.
\end{aligned}$$

The integral in the final expression factorizes over the places of $\mathbb{F}$; taking each $h_\mathfrak{p}$ to be the characteristic function of $\mathfrak{o}_p^*$ and $h_{\infty_j}(y) = h_0(Yy)$ for some fixed $h_0 \in C_c^\infty(\mathbb{R}_+^*)$ gives

$$\begin{aligned}
\mu_f(E(H, \cdot)\phi) = &\sum\sum_{\substack{(l,n) \in (\mathbb{F}^* \cap \mathfrak{o})^2 \\ m := n + l \in \mathbb{F}^* \cap \mathfrak{o}}} \frac{\lambda_\phi(l)\lambda_f(m)\lambda_f(n)}{\sqrt{|l^{\mathbf{1}} m^{\mathbf{1}} n^{\mathbf{1}}|}} \\
&\times \prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} \int_{y \in \mathbb{R}_+^*} h_0(Yy)\kappa_{\phi,\infty_j}(l_j y)\kappa_{f,\infty_j}(m_j y)\kappa_{f,\infty_j}(n_j y) \, \frac{d^\times y}{y}.
\end{aligned} \tag{2.87}$$

The integrals here, which may be treated either by bounding κ_{ϕ,∞_j} trivially as in (2.45) (which is basically what Holowinsky and Marshall do) or by our sharp refinement given in Lemma 2.4.3, essentially truncate the sum over l and n to a pair of boxes rather than regions bounded by a hyperbola and hyperplanes as in our approach.

2.6 Bounds for Shifted Sums Under Hyperbolas

In this section we establish theorem 2.4.8, whose hypotheses we now recall. Let $d = [\mathbb{F} : \mathbb{Q}]$ be the degree of our totally real number field $\mathbb{F}$, so that $\mathbb{F}_\infty \cong \mathbb{R}^d$ (see §2.2.3). Let $T \in \mathbb{R}^d_{\geq 1}$ and $U \in \mathbb{R}_{\geq 1}$ be parameters to which we associate the region

$$\mathcal{R}_{T,U} = \left\{ x \in \mathbb{R}^d : x^{\mathbf{1}} \leq X, \, x \geq T/U \right\}, \quad X := T^{\mathbf{1}}.$$

Let $\mathfrak{z} \subset \mathbb{F}$ be a fractional ideal and $l \in \mathbb{F}^* \cap \mathfrak{z}$ a nonzero "shift." Let $\lambda : I_{\mathbb{F}} \to \mathbb{C}$ be a weakly multiplicative function that satisfies $|\lambda(\mathfrak{a})| \leq \tau(\mathfrak{a})$. We would like to bound certain sums

$$\Sigma_\lambda(\mathfrak{z}, l, T, U) := \sum_{\substack{n \in \mathfrak{z} \\ m := n + l \in \mathfrak{z} \\ \mathbf{max}(m,n) \in \mathcal{R}_{T,U}}} |\lambda(\mathfrak{z}^{-1} m) \lambda(\mathfrak{z}^{-1} n)|. \tag{2.88}$$

Our strategy for doing so generalizes Holowinsky's. By the assumption $|\lambda(\mathfrak{a})| \leq \tau(\mathfrak{a})$ we reduce to quantifying the "independence" of the small prime factors of m and n, which in turn reduces to a classical sieving problem (estimating how many lattice points in a region satisfy some congruence conditions). By general machinery due to Linnik, Rényi, Bombieri and Davenport, Montgomery and others in the case $\mathbb{F} = \mathbb{Q}$ (see [7, §27], [31, p180] and [35]), such classical sieving problems follow from additive large sieve inequalities (quantifying the approximate orthogonality of a family of additive characters on a lattice when restricted to the intersection of that lattice with a sufficiently smooth region), which in turn follow from bounds for sums over well-spaced points in the support $\mathcal{R}^\wedge_{T,U}$ of the Fourier transform of a smooth majorizer for the region $\mathcal{R}_{T,U}$.

Some care is required when $[\mathbb{F} : \mathbb{Q}] > 1$ because then $\mathcal{R}^\wedge_{T,U}$ will have long and thin regions that (unfortunately) accomodate many well-spaced points. In our intended application the parameter U is small enough that one can successfully analyze $\mathcal{R}^\wedge_{T,U}$ without using any properties of $\mathfrak{z}$ beyond that it is a lattice, but to simplify our treatment and allow arbitrary values of U we instead exploit the symmetries of the fractional ideal $\mathfrak{z}$ coming from the action of the units $\mathfrak{o}^*_+$. First, we cover $\mathcal{R}_{T,U}$ by $\ll \log(eU)^{n-1}$ boxes of volume $X = T^{\mathbf{1}}$:

Lemma 2.6.1. *There exists a finite collection* $(\mathcal{R}_\alpha)_{\alpha\in A}$ *of boxes*

$$\mathcal{R}_\alpha = [a_{\alpha,1}, b_{\alpha,1}] \times \cdots \times [a_{\alpha,d}, b_{\alpha,d}] \subset \mathbb{R}^d_{\geq 0}, \quad 0 \leq a_{\alpha,j} < b_{\alpha,j}$$

whose union contains $\mathcal{R}_{T,U}$ *with* $\#A \ll \log(eU)^{d-1}$ *such that* $\mathrm{vol}(\mathcal{R}_\alpha) = X$ *and* $b_{\alpha,1}\cdots b_{\alpha,d} \ll X$ *for each* $\alpha \in A$.

Proof. Let $x \in \mathcal{R}_{T,U}$, so that $x_1 \cdots x_d \leq T_1 \cdots T_d$ and $x_i \geq T_i/U$. By the pigeonhole principle, we have $\prod_{j\neq i} x_j \leq \prod_{j\leq i} T_j$ for some index i; to simplify notation, suppose that $i = 1$, so that $x_2 \cdots x_d \leq T_2 \cdots T_d$. Choose integers $a_2, \ldots, a_d$ so that

$$\frac{T_i}{2^{a_i}} \leq x_i \leq \frac{T_i}{2^{a_i-1}}.$$

Since $0 \leq x_1 \leq T_1T_2\cdots T_d/x_2\cdots x_d \leq 2^{a_2+\cdots+a_d}T_1$, we see that x is contained in the box

$$\mathcal{R} = \left[0, 2^{a_2+\cdots+a_d}T_1\right] \times \left[\frac{T_2}{2^{a_2}}, \frac{T_2}{2^{a_2-1}},\right] \times \cdots \times \left[\frac{T_d}{2^{a_d}}, \frac{T_d}{2^{a_d-1}},\right],$$

which satisfies the desiderata of the lemma. Since $x_2 \cdots x_d \leq T_2 \cdots T_d$ implies

$$\frac{T_2}{2^{a_2}} \cdots \frac{T_d}{2^{a_d}} \leq x_2 \cdots x_d \leq T_2 \cdots T_d,$$

and because $x_i \geq T_i/U$, we deduce that

$$a_i \leq \lceil \log_2 U \rfloor \text{ for } i = 2, \ldots, d \quad \text{ and } \quad a_2 + \cdots a_d \geq 0. \tag{2.89}$$

There are $\ll \log(eU)^{d-1}$ tuples $(a_2, \ldots, a_d) \in \mathbb{Z}^{d-1}$ satisfying the conditions (2.89). $\square$

Next, because λ and $\mathfrak{z}$ are invariant under $\mathfrak{o}_+^*$, we see that for any (totally positive) unit $\eta \in \mathfrak{o}_+^*$ and any region $\mathcal{R} \subset \mathbb{R}^d$, we have

$$\sum_{\substack{n\in\mathfrak{z} \\ m:=n+l\in\mathfrak{z} \\ \mathbf{max}(m,n)\in\mathcal{R}}} |\lambda(\mathfrak{z}^{-1}m)\lambda(\mathfrak{z}^{-1}n)| = \sum_{\substack{n\in\mathfrak{z} \\ m:=n+\eta^{-1}l\in\mathfrak{z} \\ \mathbf{max}(m,n)\in\eta\mathcal{R}}} |\lambda(\mathfrak{z}^{-1}m)\lambda(\mathfrak{z}^{-1}n)|$$

where $\eta\mathcal{R} = \{\eta x : x \in \mathcal{R}\}$. The $\mathfrak{o}_+^*$-orbit of any box $\mathcal{R}_\alpha$ as in Lemma 2.6.1 contains a representative $[a_1, b_1] \times \cdots \times [a_d, b_d]$ for which $|a_i - b_i| \asymp |a_j - b_j| \asymp X^{1/d}$ for all $i, j \in \{1, \ldots, d\}$. Thus

$$\Sigma_\lambda(\mathfrak{z}, l, T, U) \ll \log(eU)^{d-1} \sup_{\mathcal{R}} \sup_{\eta\in\mathfrak{o}_+^*} \sum_{\substack{n\in\mathfrak{z} \\ m:=n+\eta^{-1}l\in\mathfrak{z} \\ \mathbf{max}(m,n)\in\mathcal{R}}} |\lambda(\mathfrak{z}^{-1}m)\lambda(\mathfrak{z}^{-1}n)| \tag{2.90}$$

where the supremum is taken over all boxes $\mathcal{R} = [a_1, b_1] \times \cdots \times [a_d, b_d]$ for which $\mathrm{vol}(\mathcal{R}) = X$, $|a_i - b_i| \asymp X^{1/d}$, $0 \leq a_i < b_i$ and $\max(b_1, \dots, b_d) \ll X^{1/d}$, with the implied constants depending only upon the field $\mathbb{F}$. Finally, if $\mathbf{max}(m, n)$ belongs to such a box $\mathcal{R}$ with $m, n \in \mathbb{F}^*_{\infty+}$, then both m and n belong to the box $(0, b_1] \times \cdots \times (0, b_d]$. Therefore theorem 2.4.8 reduces to the following result, which we shall establish in the remainder of this section.

Theorem 2.6.2. *Let $\mathbb{F}$ be a totally real number field of degree $d = [\mathbb{F} : \mathbb{Q}]$, let $\lambda : I_{\mathbb{F}} \to \mathbb{R}_{\geq 0}$ be a nonnegative-valued multiplicative function that satisfies $\lambda(\mathfrak{a}) \leq \tau(\mathfrak{a})$ for all $\mathfrak{a} \in I_{\mathbb{F}}$, let $\mathfrak{z}$ be a fractional ideal in $\mathbb{F}$, let $\lambda^0 : \mathfrak{z} \to \mathbb{R}_{\geq 0}$ be the function $\lambda^0(n) = \lambda(\mathfrak{z}^{-1}n)$, let $X \geq 2$, and let*

$$\mathcal{R}_{X,\mathfrak{z}} = (0, (\mathrm{N}(\mathfrak{z})X)^{1/d}] \times \cdots \times (0, (\mathrm{N}(\mathfrak{z})X)^{1/d}] \subset \mathbb{R}^d. \tag{2.91}$$

Then for $l \in \mathfrak{z} \cap \mathbb{F}^*$, we have

$$\sum_{\substack{n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}} \\ m := n + l \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll_{\mathbb{F},\varepsilon} \frac{X}{\log(X)^{2-\varepsilon}} \prod_{\mathrm{N}(\mathfrak{p}) \leq X} \left(1 + \frac{2\lambda(\mathfrak{p})}{\mathrm{N}(\mathfrak{p})}\right). \tag{2.92}$$

Preserve the hypotheses and notation of theorem 2.6.2. Throughout this section the nonzero shift $l \in \mathfrak{z} \cap \mathbb{F}^*$ is fixed, while m and n denote elements of $\mathfrak{z}$ having difference $m - n = l$. To ease the notation, we write $|\mathfrak{a}| = \mathrm{N}(\mathfrak{a})$ for the norm of an integral ideal $\mathfrak{a}$. Theorem 2.6.2 is trivial for bounded values of X; thus we may and shall assume for convenience that X is sufficiently large, so that for instance $\log\log(X) \gg 1$.

For a real parameter

$$z = X^{1/s}, \quad s \in \mathbb{R}_{>0}, \tag{2.93}$$

define the *z-part* of an element $n \in \mathfrak{z}$ to be the greatest divisor of the integral ideal $\mathfrak{z}^{-1}n$ each of whose prime factors has norm at most z, so that if $\mathfrak{z}^{-1}n$ factors as a product of prime powers $\prod \mathfrak{p}_i^{k_i}$, then the z-part of n is $\prod_{|\mathfrak{p}_i| \leq z} \mathfrak{p}_i^{k_i}$. Define the *z-datum* of n to be the unique triple $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c})$ of integral ideals for which

- $\mathfrak{a}$ and $\mathfrak{b}$ are coprime,
- $\mathfrak{a}\mathfrak{c}$ is the z-part of $m := n + l$, and
- $\mathfrak{b}\mathfrak{c}$ is the z-part of n.

Thus the size of $\mathfrak{c}$ quantifies the overlap between small primes occurring in $\mathfrak{z}^{-1}m$ and $\mathfrak{z}^{-1}n$. Let $\mathcal{Z}$ denote the set of all z-data that arise in this way and $\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}}$ the set of all elements $n \in \mathfrak{z}$ having z-datum $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c})$, so that we have a partition

$$\mathfrak{z} = \sqcup \{\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} : (\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}\}. \tag{2.94}$$

Note that for all $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}$ we have $\mathfrak{c}|\mathfrak{z}^{-1}l$, so that $\mathfrak{c}^{-1}\mathfrak{z}^{-1}l$ is an integral ideal.

Now let

$$y = X^{\alpha}, \quad \alpha \in \mathbb{R}_{>0} \tag{2.95}$$

be a real parameter and partition $\mathcal{Z}$ into subsets

$$\mathcal{Z}_{\leq y} = \{(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z} : \max(|\mathfrak{a}\mathfrak{c}|, |\mathfrak{b}\mathfrak{c}|) \leq y\},$$
$$\mathcal{Z}_{>y} = \{(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z} : \max(|\mathfrak{a}\mathfrak{c}|, |\mathfrak{b}\mathfrak{c}|) > y\}.$$

Thus the z-datum of $n \in \mathfrak{z}$ belongs to $\mathcal{Z}_{\leq y}$ if both $\mathfrak{z}^{-1}m$ and $\mathfrak{z}^{-1}n$ have few small prime factors and to $\mathcal{Z}_{>y}$ if either $\mathfrak{z}^{-1}m$ or $\mathfrak{z}^{-1}n$ has many small prime factors, where y determines the threshold separating "few" from "many." The latter case occurs infrequently, as we now show in Lemma 2.6.3; the former case will be addressed by Lemma 2.6.4.

Lemma 2.6.3. *Suppose that* $2 \leq z \leq y \leq X$ *with* s *and* α *as in* (2.93), (2.95) *such that* $s \asymp \log\log(X)$ *and* $\alpha \asymp 1$. *Then*

$$\sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{>y}} \sum_{\substack{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n \in \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll X \log(X)^{-A}. \tag{2.96}$$

Proof. The LHS of (2.96) is the sum of $\lambda^0(m)\lambda^0(n)$ taken over those $m, n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}$ with $m - n = l$ for which the z-part of either m or n has norm greater than y. Writing $\mathfrak{a}$ and $\mathfrak{b}$ for the z-parts of m and n and invoking Cauchy-Schwarz twice, we see that the LHS of (2.96) is

$$\leq \left(\sum_{\substack{y < |\mathfrak{a}| \leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}| \leq z}} \#(\mathfrak{a}\mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}) \right)^{1/4} \left(\sum_{m \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(m)^4 \right)^{1/4} \left(\sum_{n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(n)^2 \right)^{1/2}$$
$$+ \left(\sum_{\substack{y < |\mathfrak{b}| \leq X \\ \mathfrak{p}|\mathfrak{b} \implies |\mathfrak{p}| \leq z}} \#(\mathfrak{b}\mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}) \right)^{1/4} \left(\sum_{m \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(m)^2 \right)^{1/2} \left(\sum_{n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(n)^4 \right)^{1/4}.$$

We have $\sum_{m \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(m)^4 \ll X \log(X)^{15}$ and $\sum_{m \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}} \lambda^0(m)^2 \ll X \log(X)^3$ by the same argument as when $\mathbb{F} = \mathbb{Q}$ (see [31, §1.6]) and $\#(\mathfrak{a}\mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}) \ll 1 + |\mathfrak{a}|^{-1}X \ll |\mathfrak{a}|^{-1}X$, so that

$$\sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{>y}} \sum_{\substack{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n \in \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll X \log(X)^{O(1)} \left(\sum_{\substack{y < |\mathfrak{a}| \leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}| \leq z}} \frac{1}{|\mathfrak{a}|} \right)^{1/4}. \tag{2.97}$$

Let $\Psi(t, z)$ denote the number of integral ideals $\mathfrak{a} \subset \mathfrak{o}$ of norm $|\mathfrak{a}| \leq t$ each of whose prime divisors $\mathfrak{p}|\mathfrak{a}$ satisfy $|\mathfrak{p}| \leq z$, so that by partial summation

$$\sum_{\substack{y<|\mathfrak{a}|\leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}|\leq z}} \frac{1}{|\mathfrak{a}|} = \frac{\Psi(X,z)}{X} - \frac{\Psi(y,z)}{y} + \int_y^X \frac{\Psi(t,z)}{t^2}\, dt. \tag{2.98}$$

A theorem of Krause [38] (see also the survey [20]) asserts that

$$\Psi(t,z) = t\rho(u)\left(1 + O\left(\frac{\log(u+1)}{\log z}\right)\right), \quad u := \frac{\log t}{\log z}$$

uniformly for $t \geq 2$ and $1 \leq u \leq (\log z)^{3/5-\varepsilon}$ for any $\varepsilon > 0$, where the *Dickman function* $\rho : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfies the asymptotics $\log \rho(u) = -(1+o(1))u \log u$ as $u \to +\infty$. For $y \leq t \leq X$, our assumptions $\alpha \asymp 1$ and $s \asymp \log\log(X)$ imply that $u \asymp \log\log t$. Thus $\log z \asymp \log t/\log\log t$, so the condition for uniformity is satisfied and we obtain

$$\Psi(t,z) \ll t\exp(-2C \log\log t \log\log\log t) = t(\log t)^{-2C\log\log\log t} \ll_A t(\log t)^{-A}$$

for some $C > 0$ and every $A > 0$. It follows from (2.98) that

$$\sum_{\substack{y<|\mathfrak{a}|\leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}|\leq z}} \frac{1}{|\mathfrak{a}|} \ll_A \log(X)^{-A}. \tag{2.99}$$

We deduce the required bound by substituting (2.99) into (2.97) and taking A sufficiently large. □

On the other hand, if $\mathfrak{z}^{-1}m$ and $\mathfrak{z}^{-1}n$ have few small prime factors, then we shall show by an application of the large sieve that they typically have few *common* small prime factors; anticipating the bound given by Corollary 2.6.8, set

$$B(y,z) := \sup_{(\mathfrak{a},\mathfrak{b},\mathfrak{c})\in\mathcal{Z}_{\leq y}} \frac{\#\{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} : m,n \in \mathcal{R}_X\}}{\frac{|\mathfrak{z}^{-1}l|}{|\mathfrak{c}|^2\phi(\mathfrak{a}\mathfrak{b}\mathfrak{c}^{-1}\mathfrak{z}^{-1}l)}}, \tag{2.100}$$

where ϕ denotes the Euler phi function (multiplicative, $\mathfrak{p}^k \mapsto |\mathfrak{p}|^{k-1}(|\mathfrak{p}|-1)$).

Lemma 2.6.4. *For y, z as in* (2.93), (2.95), *we have*

$$\sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c})\in\mathcal{Z}_{\leq y}} \sum_{\substack{n\in\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n\in\mathfrak{z}\cap\mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll 4^s B(y,z)\log(X)^{\varepsilon} \prod_{|\mathfrak{p}|\leq z}\left(1 + \frac{2\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right). \tag{2.101}$$

Proof. First, write $\mathfrak{z}^{-1}m = \mathfrak{a}\mathfrak{c}\mathfrak{m}$ and factor $\mathfrak{m}$ as a product of prime powers $\mathfrak{p}_i^{a_i}$ with $|\mathfrak{p}_i| > z$;

since $|\mathfrak{m}| \leq X$, we have

$$\sum a_i \log(z) \leq \sum a_i \log|\mathfrak{p}_i| = \log|\mathfrak{m}| \leq \log(X) = s\log(z),$$

so that our assumption $\lambda(\mathfrak{p}_i^{a_i}) \leq a_i + 1 \leq 2^{a_i}$ implies $\lambda(\mathfrak{m}) \leq 2^{\sum a_i} \leq 2^s$. Writing $\mathfrak{z}^{-1}n = \mathfrak{bcn}$, we find similarly that $\lambda(\mathfrak{n}) \leq 2^s$. Since $\gcd(\mathfrak{ac}, \mathfrak{m}) = \gcd(\mathfrak{bc}, \mathfrak{n}) = \mathfrak{o}$, we obtain $\lambda^0(m)\lambda^0(n) = \lambda(\mathfrak{ac})\lambda(\mathfrak{bc})\lambda(\mathfrak{m})\lambda(\mathfrak{n}) \leq 4^s\lambda(\mathfrak{ac})\lambda(\mathfrak{bc})$. By the definition of $B(y,z)$ and the inequality $\phi(\mathfrak{ab}) \geq \phi(\mathfrak{a})\phi(\mathfrak{b})$, the LHS of (2.101) is thus

$$\leq 4^s B(y,z) \sum_{\substack{\mathfrak{c}|\mathfrak{z}^{-1}l \\ \mathfrak{p}|\mathfrak{c} \implies |\mathfrak{p}|\leq z}} \frac{|\mathfrak{z}^{-1}l|}{\phi(\mathfrak{c}^{-1}\mathfrak{z}^{-1}l)|\mathfrak{c}|^2} \sum_{\substack{|\mathfrak{ac}|\leq y \\ \mathfrak{p}|\mathfrak{ab} \implies |\mathfrak{p}|\leq z}} \sum_{|\mathfrak{bc}|\leq y} \frac{\lambda(\mathfrak{ac})\lambda(\mathfrak{bc})}{\phi(\mathfrak{a})\phi(\mathfrak{b})}. \tag{2.102}$$

For $\mathfrak{c}$ as in (2.102), the multiplicativity of λ and ϕ implies that

$$\sum_{\substack{|\mathfrak{ac}|\leq y \\ \mathfrak{p}|\mathfrak{ab} \implies |\mathfrak{p}|\leq z}} \sum_{|\mathfrak{bc}|\leq y} \frac{\lambda(\mathfrak{ac})\lambda(\mathfrak{bc})}{\phi(\mathfrak{a})\phi(\mathfrak{b})} \leq \left(\prod_{|\mathfrak{p}|\leq z} \sum_{k\geq 0} \frac{\lambda(\mathfrak{p}^{k+v_{\mathfrak{p}}(\mathfrak{c})})}{\phi(\mathfrak{p}^k)} \right)^2, \tag{2.103}$$

where $v_{\mathfrak{p}}(\mathfrak{c})$ denotes the order to which $\mathfrak{p}$ divides $\mathfrak{c}$. We rewrite

$$\sum_{k\geq 0} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} = \left(1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right) \left(1 + \frac{\frac{\lambda(\mathfrak{p})}{\phi(\mathfrak{p})} - \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|} + \sum_{k\geq 2} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)}}{1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}}\right). \tag{2.104}$$

Using the inequalities $\lambda(\mathfrak{p}^k) \leq k+1$ and $|\mathfrak{p}| \geq 2$ and writing $q = |\mathfrak{p}|$ for clarity, we compute

$$\begin{aligned} \frac{\lambda(\mathfrak{p})}{\phi(\mathfrak{p})} - \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|} + \sum_{k\geq 2} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} &\leq \frac{2}{q(q-1)} + \sum_{k\geq 2} \frac{k+1}{q^{k-1}(q-1)} \\ &= q^{-2}\left(2(1-q^{-1})^{-1} + 2(1-q^{-1})^{-2} + (1-q^{-1})^{-3}\right) \\ &\leq 20q^{-2}, \end{aligned}$$

so that (2.104) implies

$$\sum_{k\geq 0} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} \leq \left(1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right)\left(1 + \frac{20}{|\mathfrak{p}|^2}\right). \tag{2.105}$$

If $\nu \geq 1$, then (writing $q = |\mathfrak{p}|$)

$$\begin{aligned} \sum_{k\geq 0} \frac{\lambda(\mathfrak{p}^{k+\nu})}{\phi(\mathfrak{p}^k)} &\leq \nu + 1 + \sum_{k\geq 1} \frac{\nu+k+1}{q^{k-1}(q-1)} \\ &= 1 + \nu\left(1 + q^{-1}(1-q^{-1})^{-2}\right) + q^{-1}(1-q^{-1})^{-2} \\ &\leq 3\nu + 3. \end{aligned}$$

Substituting these bounds into (2.102) and (2.103), the LHS of (2.101) is

$$\begin{aligned} &\ll 4^s B(y,z)\psi(\mathfrak{z}^{-1}l)\prod_{|\mathfrak{p}|\le z}\left(1+\frac{2\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right), \\ &\text{with } \psi(\mathfrak{a}) := |\mathfrak{a}|\sum_{\mathfrak{c}|\mathfrak{a}}\frac{\prod_{\mathfrak{p}^\nu||\mathfrak{c}}(3\nu+3)^2}{\phi(\mathfrak{a}/\mathfrak{c})|\mathfrak{c}|^2}. \end{aligned} \tag{2.106}$$

The function $\psi : I_{\mathbb{F}} \to \mathbb{R}_{\ge 0}$ is multiplicative. On a prime power $\mathfrak{p}^a$ with $a \ge 1$ and $|\mathfrak{p}| = q \ge 2$ it takes the value

$$\psi(\mathfrak{p}^k) = \frac{1}{1-q^{-1}} + \frac{9}{q^a}\left((a+1)^2 + \frac{1}{1-q^{-1}}\sum_{i=1}^{a-1}\frac{(i+1)^2}{q^i}\right) \le 1 + 10^6 q^{-1}.$$

Since $\prod_{\mathfrak{p}|\mathfrak{a}}(1+|\mathfrak{p}|^{-1}) \ll \log\log|\mathfrak{a}|$, it follows that $\psi(\mathfrak{a}) \ll \log\log(\mathfrak{a})^{10^6}$. If $|\mathfrak{z}^{-1}l| > X$, then the LHS of (2.101) is zero; if otherwise $|\mathfrak{z}^{-1}l| \le X$, then $\psi(\mathfrak{z}^{-1}\mathfrak{l}) \ll \log(X)^{\varepsilon}$. Thus (2.101) follows from (2.106). $\square$

By Lemma 2.6.3 and Lemma 2.6.4, we see that theorem 2.4.8 follows from sufficiently strong bounds for the quantity $B(y,z)$ given by (2.100); the following lemma reduces such bounds to a classical sieving problem.

Definition 2.6.5. For a region $\mathcal{R} \subset \mathbb{F}_\infty \cong \mathbb{R}^d$, an ideal $\mathfrak{x} \subset \mathbb{F}$, a finite set $\mathcal{P}$ of primes in $\mathfrak{o}$ and a collection $(\Omega_\mathfrak{p})_{\mathfrak{p}\in\mathcal{P}}$ of sets of residue classes $\Omega_\mathfrak{p} \subset \mathfrak{x}/\mathfrak{p}\mathfrak{x}$, define the *sifted set*

$$\mathcal{S}(\mathcal{R},\mathfrak{x},(\Omega_\mathfrak{p})) := \{n \in \mathfrak{x}\cap\mathcal{R} : n \notin \Omega_\mathfrak{p} \quad (\mathfrak{p}\mathfrak{x}) \text{ for all } \mathfrak{p}\in\mathcal{P}\}. \tag{2.107}$$

Define also for any $Q \ge 1$ the quantity

$$H((\Omega_\mathfrak{p}),Q) = \sum_{\substack{|\mathfrak{q}|\le Q \\ \mathfrak{p}|\mathfrak{q} \implies \mathfrak{p}\in\mathcal{P}}}\prod_{\mathfrak{p}|\mathfrak{q}}\frac{\#\Omega_\mathfrak{p}}{|\mathfrak{p}|-\#\Omega_\mathfrak{p}}. \tag{2.108}$$

Lemma 2.6.6. *Let $(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}$. Choose an element $r \in \mathfrak{c}\mathfrak{z}$ so that $r \equiv 0$ $(\mathfrak{a}\mathfrak{c}\mathfrak{z})$ and $r = -l$ $(\mathfrak{b}\mathfrak{c}\mathfrak{z})$, and define the region*

$$\mathcal{R}_r = \{x - r | x \in \mathcal{R}_{X,\mathfrak{z}}\}. \tag{2.109}$$

Let $\mathfrak{x} = \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ and let $\mathcal{P}$ denote the set of odd primes $\mathfrak{p}$ in $\mathfrak{o}$ of norm $|\mathfrak{p}| \le z$. Then there exists a collection of sets of residue classes $(\Omega_\mathfrak{p})_{\mathfrak{p}\in\mathcal{P}}$ with $\Omega_\mathfrak{p} \subset \mathfrak{x}/\mathfrak{p}\mathfrak{x}$ such that

$$\#\Omega_\mathfrak{p} := \begin{cases} 1 & \mathfrak{p}|\mathfrak{a}\mathfrak{b}\mathfrak{c}^{-1}\mathfrak{z}^{-1}l \\ 2 & \textit{otherwise} \end{cases} \tag{2.110}$$

and

$$\#(\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \cap \mathcal{R}_{X,\mathfrak{z}}) \leq \#\mathcal{S}(\mathcal{R}_r, \mathfrak{x}, (\Omega_\mathfrak{p})). \tag{2.111}$$

Proof. Indeed, let $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}$, so that $\mathfrak{c}|\mathfrak{z}^{-1}l$ and $\gcd(\mathfrak{a}, \mathfrak{b}) = \mathfrak{o}$. Let $n \in \mathfrak{z}$. Then n belongs to $\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}}$ if and only if

(1) $n \in \mathfrak{a}\mathfrak{c}\mathfrak{z}$,

(2) $n + l \in \mathfrak{b}\mathfrak{c}\mathfrak{z}$,

(3) $\mathfrak{p} \nmid \mathfrak{z}^{-1}n/\mathfrak{a}\mathfrak{c}$ for each prime $\mathfrak{p}$ with norm $|\mathfrak{p}| \leq z$, and

(4) $\mathfrak{p} \nmid \mathfrak{z}^{-1}(n + l)/\mathfrak{b}\mathfrak{c}$ for each prime $\mathfrak{p}$ with norm $|\mathfrak{p}| \leq z$.

If $n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}}$, then conditions (1)–(2) assert that $n - r \in \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$, while conditions (3)–(4) assert (slightly more than) that for each prime $\mathfrak{p}$ with $|\mathfrak{p}| \leq z$, the number $n - r \in \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ does not belong to a certain collection $\Omega_\mathfrak{p} \subset \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ of residue classes. Precisely, let $\zeta \in \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ and $n = \zeta + r$.

- Suppose $\mathfrak{p}|\mathfrak{a}$, $\mathfrak{p} \nmid \mathfrak{b}$. Let $\zeta_1 := (\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z} \xrightarrow{\cong} \mathfrak{a}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{c}\mathfrak{z})^{-1}(-r)$. Then (3) holds iff $\zeta + r \notin \mathfrak{p}\mathfrak{a}\mathfrak{c}\mathfrak{z}$ iff $\zeta - \zeta_1 \notin \mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$, while (4) holds iff $\zeta + r + l \notin \mathfrak{p}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ iff (since $\zeta \in \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z} \subset \mathfrak{p}\mathfrak{b}\mathfrak{c}\mathfrak{z}$) $r + l \notin \mathfrak{p}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ iff $\mathfrak{p}\mathfrak{b}\mathfrak{z} \nmid \frac{r+l}{\mathfrak{c}}$ iff (since $(\mathfrak{p}, \mathfrak{b}) = 1$ and $r + l \in \mathfrak{b}\mathfrak{c}$) $r + l \notin \mathfrak{p}\mathfrak{c}\mathfrak{z}$; we may take $\Omega_\mathfrak{p} = \{\zeta_1\}$, $\#\Omega_\mathfrak{p} = 1$.
- If $\mathfrak{p} \nmid \mathfrak{a}$, $\mathfrak{p}|\mathfrak{b}$, then we may similarly take $\#\Omega_\mathfrak{p} = 1$.
- The case $\mathfrak{p}|\mathfrak{a}, \mathfrak{p}|\mathfrak{b}$ does not occur because $(\mathfrak{a}, \mathfrak{b}) = 1$.
- Suppose $\mathfrak{p} \nmid \mathfrak{a}\mathfrak{b}$. Let $\zeta_1 := (\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z} \xrightarrow{\cong} \mathfrak{a}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{c}\mathfrak{z})^{-1}(-r)$, $\zeta_2 := (\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z} \xrightarrow{\cong} \mathfrak{b}\mathfrak{c}\mathfrak{z}/\mathfrak{p}\mathfrak{b}\mathfrak{c}\mathfrak{z})^{-1}(-r - l)$. Then (3) holds iff $\zeta + r \notin \mathfrak{p}\mathfrak{a}\mathfrak{c}\mathfrak{z}$ iff $\zeta - \zeta_1 \notin \mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$, while (4) holds iff $\zeta + r + l \notin \mathfrak{p}\mathfrak{b}\mathfrak{c}\mathfrak{z}$ iff $\zeta - \zeta_2 \notin \mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}$. We may therefore take $\Omega_\mathfrak{p} = \{\zeta_1, \zeta_2\}$. We have $\zeta_1 \equiv \zeta_2$ $(\mathfrak{p}\mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z})$ iff $l \in \mathfrak{p}\mathfrak{c}\mathfrak{z}$, in which case $\#\Omega_\mathfrak{p} = 1$; if $l \notin \mathfrak{p}\mathfrak{c}\mathfrak{z}$, then $\#\Omega_\mathfrak{p} = 2$.

Thus $n \mapsto n-r$ gives an inclusion $\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \cap \mathcal{R} \hookrightarrow \mathcal{S}(\mathcal{R}_r, \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{z}, (\Omega_\mathfrak{p}))$, and the $\#\Omega_\mathfrak{p}$ are as claimed. □

The large sieve machinery alluded to above allows us to show the following, the proof of which we postpone to a later subsection; the proof is independent of what follows in this subsection, so there is no circularity in our arguments.

Proposition 2.6.7. *Let $\mathfrak{x}$, $\mathcal{P}$, and $(\Omega_\mathfrak{p})_{\mathfrak{p}\in\mathcal{P}}$ be as in Definition 2.6.5. Let $\mathcal{R}$ be the region $\mathcal{R}_{X,\mathfrak{x}}$ as in* (2.91) *or a translate thereof. There exists a positive constant $c_2(\mathbb{F}) > 0$ such that for $X > c_2(\mathbb{F})$ and $Q \geq 1$, we have*

$$\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_\mathfrak{p})) \ll \frac{X + Q^2}{H((\Omega_\mathfrak{p}), Q)}. \tag{2.112}$$

Proof. See §2.7. ☐

As a consequence, we deduce the following bound for $B(y, z)$.

Corollary 2.6.8. *Let $c_2(\mathbb{F}) > 0$ be as in Proposition 2.6.7. Then for $X > c_2(\mathbb{F})y^2$, the quantity $B(y, z)$ given by* (2.100) *satisfies*

$$B(y, z) \ll \frac{X + y^2z^2}{\log(z)^2}.$$

Proof. Let $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}_{\leq y}$ and let the region $\mathcal{R}_r$, the ideal $\mathfrak{x} = \mathfrak{abcz}$, the set of primes $\mathcal{P}$ and the collection of sets of residue classes $(\Omega_\mathfrak{p})$ be as in Lemma 2.6.6, so that (2.111) holds. Then $|\mathfrak{x}| \leq y^2|\mathfrak{z}|$, so that $X > c_2(\mathbb{F})y^2$ implies $X' > c_2(\mathbb{F})$ with $X' := |\mathfrak{x}^{-1}\mathfrak{z}|X$; the hypothesis of Proposition 2.6.7 are then satisfied (taking X' in place of X), and setting $Q = z$ we obtain

$$\#(\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \cap \mathcal{R}_{X,\mathfrak{z}}) \ll \frac{|\mathfrak{x}^{-1}\mathfrak{z}|X + z^2}{H((\Omega_\mathfrak{p}), z)}.$$

Set $\mathfrak{m} = \mathfrak{abc}^{-1}\mathfrak{z}^{-1}l$ (see (2.110)). The lower bound

$$H((\Omega_\mathfrak{p}), z) \gg_\mathbb{F} \frac{\phi(\mathfrak{m})}{|\mathfrak{m}|} \log(z)^2$$

is standard when $\mathbb{F} = \mathbb{Q}$ and follows in general from the arguments of [17, pp55-59, Thm 2] upon redefining "$P(z)$" to be the product of all prime ideals of norm up to z, replacing every sum over integers (resp. primes) satisfying some inequalities by the analogous sum over ideals (resp. prime ideals) with norms satisfying the analogous inequalities, and replacing the Riemann zeta function ζ by the Dedekind zeta function $\zeta_\mathbb{F}$. Thus recalling the definition (2.100) of $B(y, z)$, we obtain

$$B(y, z) \ll \frac{|\mathfrak{x}|^{-1}X + z^2}{\frac{\phi(\mathfrak{m})}{|\mathfrak{m}|}\log(z)^2} \frac{|\mathfrak{c}|^2\phi(\mathfrak{m})}{|\mathfrak{z}^{-1}l|} = \frac{X + |\mathfrak{abc}|z^2}{\log(z)^2}.$$

Since $|\mathfrak{abc}| \leq y^2$, we deduce the claimed bound. ☐

Proof of theorem 2.6.2. Let y, z be given by (2.93), (2.95) with $\alpha \in (0, \frac{1}{2})$ and $s = \alpha \log\log(X)$. We eventually (i.e., as $X \to \infty$) have $X > c_2(\mathbb{F})y^2$ and $2 \leq z \leq y \leq X$. Thus the hypotheses of Lemma 2.6.3, Lemma 2.6.4 and Corollary 2.6.8 are eventually satisfied, so we obtain

$$\sum_{\substack{n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}} \\ m := n + l \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll 4^s \frac{X + y^2z^2}{\log(z)^2} \log(X)^\varepsilon \prod_{\mathrm{N}(\mathfrak{p}) \leq z} \left(1 + \frac{2\lambda(\mathfrak{p})}{\mathrm{N}(\mathfrak{p})}\right).$$

We have $4^s = \log(X)^{\alpha \log(4)}$, $\log(z) \gg_\alpha \log(X)^{2-\varepsilon}$ and $y^2z^2 \ll_\alpha X$, so letting $\alpha \to 0$ we deduce the assertion of theorem 2.6.2. ☐

2.7 Appendix: Sieve Bounds

Inequalities of the shape (2.112) (with explicit constants) have appeared in papers of Schaal [55, Thm 5] and Hinz [21, Satz 2], but only under additional assumptions such as $Q \gg_{\mathbb{F}} 1$, $X \gg Q^2$, and $\Omega_{\mathfrak{p}} = \emptyset$ for all $\mathfrak{p}|\mathfrak{z}$. Although it would possible to get around such assumptions in our intended applications (at the cost of sacrificing the uniformity in $\mathfrak{z}$, which is ultimately not needed), we prefer to establish a result in which such assumptions are not present. We neglect here the issue of the leading coefficient of such bounds, which is important in some of the applications of the authors just cited but not in ours; for this reason our analysis is substantially simplified.

Our arguments in this short section are standard; we have been influenced by the books of Davenport [7] and Kowalski [35], to which we refer the reader for a discussion of the history of these ideas. Fix a fractional ideal $\mathfrak{x}$ of $\mathbb{F}$. Let $\mathfrak{q}$ be an integral ideal in $\mathbb{F}$ and $\alpha : \mathfrak{x}/\mathfrak{q}\mathfrak{x} \to \mathbb{C}$ a function on the group $\mathfrak{x}/\mathfrak{q}\mathfrak{x}$. Define $L^2(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$, $\|.\|_2$ with respect to the counting measure, and for ψ in the Pontryagin dual $(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge$, define $\alpha^\wedge(\psi) = \sum_{\mathfrak{x}/\mathfrak{q}\mathfrak{x}} \alpha(\zeta)\overline{\psi}(\zeta)$; then the Fourier inversion and Plancherel formulas read

$$\alpha = |\mathfrak{q}|^{-1} \sum_{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge} \alpha^\wedge(\psi)\psi, \quad \sum_{\mathfrak{x}/\mathfrak{q}\mathfrak{x}} |\alpha(\zeta)|^2 = \|\alpha\|_2^2 = \|\alpha^\wedge\|_2^2 = |\mathfrak{q}|^{-1} \sum_{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge} |\alpha^\wedge(\psi)|^2.$$

For a proper divisor $\mathfrak{q}'$ of $\mathfrak{q}$, the projection $\mathfrak{x}/\mathfrak{q}\mathfrak{x} \to \mathfrak{x}/\mathfrak{q}'\mathfrak{x}$ induces an inclusion $L^2(\mathfrak{x}/\mathfrak{q}'\mathfrak{x}) \hookrightarrow L^2(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$. Let $L^2_{\#}(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$ denote the orthogonal complement of the span of the images of these inclusions, write $L^2(\mathfrak{x}/\mathfrak{q}\mathfrak{x}) \ni \alpha \mapsto \alpha_{\#} \in L^2_{\#}(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$ for the associated orthogonal projection, and let $(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_{\#}$ denote the set of characters $\psi \in (\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge$ that do not factor through any proper projection $\mathfrak{x}/\mathfrak{q}\mathfrak{x} \to \mathfrak{x}/\mathfrak{q}'\mathfrak{x}$, so that

$$\|\alpha_{\#}\|_2^2 = |\mathfrak{q}|^{-1} \sum_{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_{\#}} |\alpha^\wedge(\psi)|^2.$$

For $\psi \in (\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_{\#}$ call $\mathfrak{q}$ the *conductor* of ψ.

Let $\mathcal{R}$ be a region in $\mathbb{F}_\infty$, $\mathcal{P}$ a finite set of primes, $Q \geq 1$ a parameter, and $\mathcal{Q}$ the set of squarefree ideals $\mathfrak{q}$ composed of primes $\mathfrak{p} \in \mathcal{P}$ with $|\mathfrak{q}| \leq Q$. Let $V(\mathcal{R}, \mathfrak{x})$ be the Hilbert space of complex-valued functions $(a_n)_n : \mathfrak{x} \to \mathbb{C}$ supported on $\mathcal{R} \cap \mathfrak{x}$, where for $(a_n) \in V(\mathcal{R}, \mathfrak{x})$ we set $\|a\|_2^2 := \sum_n |a_n|^2$. For $\mathfrak{q} \in \mathcal{Q}$ define $a[\mathfrak{q}] \in L^2(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$ by the formula $a[\mathfrak{q}](\zeta) = \sum_{n=\zeta(\mathfrak{q}\mathfrak{x})} a_n$. Let $E(\cdot; \mathfrak{x}, Q)$ be the quadratic form on $V(\mathcal{R}, \mathfrak{x})$ defined by

$$E((a_n); \mathfrak{x}, Q) = \sum_{\mathfrak{q} \in \mathcal{Q}} |\mathfrak{q}| \|a[\mathfrak{q}]_{\#}\|_2^2 = \sum_{\mathfrak{q} \in \mathcal{Q}} \sum_{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_{\#}} |a[\mathfrak{q}]^\wedge(\psi)|^2, \tag{2.113}$$

and $D(\mathcal{R}, \mathfrak{x}, Q)$ the squared norm of $E(\cdot; \mathfrak{x}, Q)$, i.e., the smallest non-negative real with the property that $|E((a_n); \mathfrak{x}, Q)| \leq D(\mathcal{R}, \mathfrak{x}, Q)\|a\|_2^2$ for all $(a_n) \in V(\mathcal{R}, \mathfrak{x})$.

Suppose that $\alpha[\mathfrak{p}](\zeta) = 0$ for (at least) $\omega(\mathfrak{p})$ values of ζ mod $\mathfrak{p}$ for each $\mathfrak{p} \in \mathcal{P}$, and set $h(\mathfrak{q}) = \prod_{\mathfrak{p}|\mathfrak{q}} \frac{\omega(\mathfrak{p})}{|\mathfrak{p}|-\omega(\mathfrak{p})}$ for each $\mathfrak{q} \in \mathcal{Q}$. An inequality due to Montgomery [44] in the $(\mathbb{F}, \mathfrak{x}) = (\mathbb{Q}, \mathbb{Z})$ case (refining earlier work of Linnik, Rényi, and Bombieri-Davenport), whose proof generalizes painlessly to the present situation and has been formulated axiomatically by Kowalski [35, Lem 2.7], shows that $h(\mathfrak{q})\|a[\mathfrak{o}]\|_2^2 \leq |\mathfrak{q}|\|a[\mathfrak{q}]_{\#}\|_2^2$, so recalling from (2.108) that $H((\Omega_{\mathfrak{p}}), Q) = \sum_{\mathfrak{q} \in \mathcal{Q}} h(\mathfrak{q})$ we obtain

$$\|a[\mathfrak{o}]\|_2^2 H((\Omega_{\mathfrak{p}}), Q) \leq D(\mathcal{R}, \mathfrak{x}, Q)\|a\|_2^2.$$

In the special case that $(a_n)_n$ is the indicator function of $\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_{\mathfrak{p}}))$ for some subsets $\Omega_{\mathfrak{p}} \subset \mathfrak{x}/\mathfrak{p}\mathfrak{x}$, let $Z := \#\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_{\mathfrak{p}}))$, so that

$$\|a\|_2^2 = \sum_n |a_n|^2 = Z, \quad \|a[\mathfrak{o}]\|_2^2 = |\sum_n a_n|^2 = Z^2,$$

and $a_n = 0$ whenever $n \in \Omega_{\mathfrak{p}}\ \ (\mathfrak{p})$ for any $\mathfrak{p} \in \mathcal{P}$. Thus

$$\#\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_{\mathfrak{p}})) \leq \frac{D(\mathcal{R}, \mathfrak{x}, Q)}{H((\Omega_{\mathfrak{p}}), Q)}. \tag{2.114}$$

In this context, an additive large sieve inequality is by definition a bound for $D(\mathcal{R}, \mathfrak{x}, Q)$. The homomorphism $\mathbb{F}_\infty/\mathfrak{x}^{-1}\mathfrak{d}^{-1} \ni \xi \mapsto [\mathfrak{x} \ni n \mapsto e(\operatorname{Tr} \xi n)] \in \mathfrak{x}^\wedge$ $(e(x) = e^{2\pi i x})$ induces for integral ideals $\mathfrak{q}'|\mathfrak{q}$ the compatible isomorphisms

$$\begin{array}{ccc}
\mathfrak{q}'^{-1}\mathfrak{x}^{-1}\mathfrak{d}^{-1}/\mathfrak{x}^{-1}\mathfrak{d}^{-1} & \xrightarrow{\cong} & (\mathfrak{x}/\mathfrak{q}'\mathfrak{x})^\wedge \\
\downarrow & & \downarrow \\
\mathfrak{q}^{-1}\mathfrak{x}^{-1}\mathfrak{d}^{-1}/\mathfrak{x}^{-1}\mathfrak{d}^{-1} & \xrightarrow{\cong} & (\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge
\end{array}$$

by which we regard the family $\sqcup\{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_\# : \mathfrak{q} \in \mathcal{Q}\}$ of primitive additive characters having (squarefree) conductor up to Q (and supported on the primes of $\mathcal{P}$) as a subset $\mathcal{F} := \mathcal{F}(\mathfrak{x}, Q) \subset \mathbb{F}/\mathfrak{x}^{-1}\mathfrak{d}^{-1} \subset \mathbb{F}_\infty/\mathfrak{x}^{-1}\mathfrak{d}^{-1}$ of the family of all (finite order) additive characters on $\mathfrak{x}$, thus

$$E((a_n); \mathfrak{x}, Q) = \sum_{\xi \in \mathcal{F}(\mathfrak{x}, Q)} \left| \sum_n a_n e(\operatorname{Tr} \xi n) \right|^2. \tag{2.115}$$

Write $D(\mathcal{R}, \mathfrak{x}, \mathcal{F})$ synonymously for $D(\mathcal{R}, \mathfrak{x}, Q)$. The group $\mathfrak{o}_+^*$ acts on $\mathbb{F}_\infty$ and $\mathbb{F}_\infty/\mathfrak{x}^{-1}\mathfrak{d}^{-1}$ by multiplication, stabilizing $\mathfrak{x}$ and $\mathcal{F}$. The ℓ^∞ metric on $\mathbb{F}_\infty$ given by $d_{\mathbb{F}_\infty}(\xi, \eta) = \max_i |\xi_i - \eta_i|$ induces on $\mathbb{F}_\infty/\mathfrak{x}^{-1}\mathfrak{d}^{-1}$ by the formula $d(\xi, \eta) := \min_{n \in \mathfrak{x}^{-1}\mathfrak{d}^{-1}} d_{\mathbb{F}_\infty}(\xi, \eta + n)$ a metric d with respect to which we call

$$\delta := \delta(\mathcal{F}(\mathfrak{r}, Q)) := \min_{\xi \neq \eta \in \mathcal{F}(\mathfrak{r}, Q)} d(\xi, \eta)$$

the smallest spacing for the family $\mathcal{F}(\mathfrak{r}, Q)$ and say that $\mathcal{F}(\mathfrak{r}, Q)$ is $\delta(\mathcal{F}(\mathfrak{r}, Q))$*-spaced.*

Lemma 2.7.1. $\delta(\mathcal{F}(\mathfrak{r}, Q)) \geq (|\mathfrak{r}|\Delta_{\mathbb{F}} Q^2)^{-1/[\mathbb{F}:\mathbb{Q}]}$ *(here* $\Delta_{\mathbb{F}} = |\mathfrak{d}|$ *is the discriminant of* $\mathbb{F}$*).*

Proof. Suppose that $\mathfrak{q}_1, \mathfrak{q}_2 \in \mathcal{Q}$, $\xi \in \mathfrak{q}_1^{-1}\mathfrak{r}^{-1}\mathfrak{d}^{-1}$, and $\eta \in \mathfrak{q}_2^{-1}\mathfrak{r}^{-1}\mathfrak{d}^{-1}$ with $\xi - \eta \notin \mathfrak{r}^{-1}\mathfrak{d}^{-1}$. We must show, for any $n \in \mathfrak{r}^{-1}\mathfrak{d}^{-1}$, that $\zeta := \xi - \eta - n$ satisfies $\max_i |\zeta_i| \geq (|\mathfrak{r}|\Delta_{\mathbb{F}} Q^2)^{-1/[\mathbb{F}:\mathbb{Q}]}$. Indeed, we have $0 \neq \zeta \in \mathfrak{q}_1^{-1}\mathfrak{q}_2^{-1}\mathfrak{r}^{-1}\mathfrak{d}^{-1}$, so that

$$\prod |\xi_i - \eta_i| = |\xi - \eta|^{\mathbf{1}} \geq |\mathfrak{q}_1^{-1}\mathfrak{q}_2^{-1}\mathfrak{r}^{-1}\mathfrak{d}^{-1}| \geq \Delta_{\mathbb{F}}^{-1}|\mathfrak{r}|^{-1}Q^{-2}.$$

Thus for some index i we have $|\zeta_i| \geq (|\mathfrak{r}|\Delta_{\mathbb{F}} Q^2)^{-1/[\mathbb{F}:\mathbb{Q}]}$, hence the claim. □

The duality principle for bilinear forms, which asserts that a form and its transpose have the same norm, implies that $D(\mathcal{R}, \mathfrak{r}, \mathcal{F})$ is the smallest non-negative real such that

$$\sum_{n \in \mathfrak{r} \cap \mathcal{R}} \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2 \leq D(\mathcal{R}, \mathfrak{r}, \mathcal{F}) \|b\|_2^2 \tag{2.116}$$

for all $(b_\xi)_\xi : \mathcal{F} \to \mathbb{C}$, where $\|b\|_2^2 = \sum |b_\xi|^2$. Call a nonnegative-valued Schwarz function $f \in \mathcal{S}(\mathbb{F}_\infty \to \mathbb{R}_{\geq 0})$ $\mathcal{R}$*-admissible* if it satisfies $f|_{\mathcal{R}} \geq 1$, and let f be $\mathcal{R}$-admissible. Opening the square in (2.116) and invoking the elementary inequality $|b_\xi \overline{b_\eta}| \leq \frac{1}{2}(|b_\xi|^2 + |b_\eta|^2)$, we find that

$$\begin{aligned} \sum_{n \in \mathfrak{r} \cap \mathcal{R}} \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2 &\leq \sum_{n \in \mathfrak{r}} f(n) \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2 \\ &\leq \sup_{\xi \in \mathcal{F}} \sum_{\eta \in \mathcal{F}} \left| \sum_{n \in \mathfrak{r}} f(n) e(\operatorname{Tr} n(\xi - \eta)) \right| \|b\|_2^2. \end{aligned}$$

Applying the Poisson summation formula, which asserts in this context that

$$\begin{aligned} \sum_{n \in \mathfrak{r}} f(n) e(\operatorname{Tr} n(\xi - \eta)) = \operatorname{vol}(\mathbb{F}_\infty/\mathfrak{r})^{-1} &\sum_{\mu \in \mathfrak{r}^{-1}\mathfrak{d}^{-1}} \hat{f}(\mu - \xi + \eta), \\ \text{with } \hat{f}(y) := &\int_{\mathbb{F}_\infty} f(x) e(-x \cdot y)\, dy, \end{aligned}$$

we obtain

$$D(\mathcal{R},\mathfrak{r},\mathcal{F}) \leq \mathrm{vol}(\mathbb{F}_\infty/\mathfrak{o})^{-1}|\mathfrak{r}|^{-1}F(f;\mathfrak{r},\mathcal{F}),$$
$$\text{with } F(f;\mathfrak{r},\mathcal{F}) := \sup_{\xi\in\mathcal{F}}\sum_{\eta\in\mathcal{F}}\left|\sum_{\mu\in\mathfrak{r}^{-1}\mathfrak{d}^{-1}}\hat{f}(\mu-\xi+\eta)\right|. \tag{2.117}$$

Lemma 2.7.2. *There exists a positive constant $c_2(\mathbb{F}) > 0$ with the following property. For any rectangle $\mathcal{R} = \prod[a_i,b_i] = [a_1,b_1]\times\cdots\times[a_d,b_d]$ whose volume $\mathrm{vol}(\mathcal{R}) = \prod|a_i-b_i|$ satisfies $\mathrm{vol}(\mathcal{R}) > c_2(\mathbb{F})|\mathfrak{r}|$, there exists an $\mathcal{R}$-admissible function f such that*

$$F(f,\mathfrak{r},\mathcal{F}) \ll_{\mathbb{F}} \mathrm{vol}(\mathcal{R}) + \delta^{-d}. \tag{2.118}$$

Proof. For a unit $\eta\in\mathfrak{o}_+^*$ and an $\mathcal{R}$-admissible function f, define the $\eta\mathcal{R}$-admissible function ηf by the formula $\eta f(\eta x) = f(x)$. Since $\mathfrak{r}$ and $\mathcal{F}$ are $\mathfrak{o}_+^*$-stable, we have $F(\eta f;\mathfrak{r},\mathcal{F}) = F(f;\mathfrak{r},\mathcal{F})$. Therefore we may assume that $\mathcal{R}$ is chosen so that $|a_i-b_i|\asymp|a_j-b_j|$ for all $i,j\in\{1,\dots,d\}$, where the implied constant depends only upon $\mathbb{F}$. Now the formula

$$f(x) = \left(\frac{\pi^2}{8}\right)^d\prod_{i=1}^d \mathrm{sinc}^2\left(\frac{x_i-\frac{a_i+b_i}{2}}{2|a_i-b_i|}\right), \quad \mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

defines an $\mathcal{R}$-admissible function f whose Fourier transform is supported in the *dual rectangle*

$$\widehat{\mathcal{R}} = \prod[c_i,d_i], \quad |c_i-d_i| = |a_i-b_i|^{-1}, \quad c_i = -d_i < 0 < d_i$$

and satisfies $\|\hat{f}\|_\infty \leq (\pi^2/4)^d\prod|a_i-b_i|$. Since $|a_i-b_i|\asymp|a_j-b_j|$ for all i,j, there exists a constant $c_2(\mathbb{F}) > 0$, depending only upon $\mathbb{F}$, such that $\mathrm{vol}(\mathcal{R}) > c_2(\mathbb{F})|\mathfrak{r}|$ implies that $|a_i-b_i| > \frac{1}{2}\Delta_{\mathbb{F}}^{1/d}|\mathfrak{r}|^{1/d}$ for each i. If we assume now (as we may) that the latter assertion holds, then any translate of the dual rectangle $\widehat{\mathcal{R}}$ contains at most one element of the dual lattice $\mathfrak{r}^{-1}\mathfrak{d}^{-1}$, so that each sum over μ in (2.117) contains at most one nonzero term, thus

$$\sum_{\eta\in\mathcal{F}}\left|\sum_{\mu\in\mathfrak{r}^{-1}\mathfrak{d}^{-1}}\hat{f}(\mu-\xi+\eta)\right| \leq \|\hat{f}\|_\infty\cdot\#\left\{\eta\in\mathcal{F} : \mu-\xi+\eta\in\widehat{\mathcal{R}}+\mathfrak{r}^{-1}\mathfrak{d}^{-1}\right\}.$$

The above set is a δ-spaced subset of $\widehat{\mathcal{R}}$ (mod $\mathfrak{r}^{-1}\mathfrak{d}^{-1}$); a cube-packing argument shows that any such set has cardinality at most $\prod(1+\lfloor\delta^{-1}|c_i-d_i|\rfloor)$, so that

$$F(f,\mathfrak{r},\mathcal{F}) \leq \left(\frac{\pi^2}{4}\right)^d\prod_{i=1}^d|a_i-b_i|(1+\lfloor\delta^{-1}|c_i-d_i|\rfloor) \ll \prod_{i=1}^d(|a_i-b_i|+\delta^{-1}). \tag{2.119}$$

Since $|a_i - b_i| \asymp |a_j - b_j|$, we obtain $F(f, \mathfrak{x}, \mathcal{F}) \ll \mathrm{vol}(\mathcal{R}) + \delta^{-d}$, as desired. $\square$

Proof of Proposition 2.6.7. Take $c_2(\mathbb{F})$ as in Lemma 2.7.2, and suppose that $X > c_2(\mathbb{F})$ and $Q \geq 1$. Then $\mathrm{vol}(\mathcal{R}_{X,\mathfrak{z}}) > c_2(\mathbb{F})|\mathfrak{z}|$, so the hypotheses of Lemma 2.7.2 are satisfied. The claimed bound (2.112) follows immediately from (2.114), Lemma 2.7.1, equation (2.117) and Lemma 2.7.2. $\square$

2.8 Appendix: Bounds for Special Functions

In this self-contained section we establish the technical lemmas that were needed in the proof of Lemma 2.4.3. First, recall [72] that the Gauss hypergeometric function $F = {}_2F_1$ is defined for $\mathrm{Re}(c) > \mathrm{Re}(b) > 0$ and $|\arg(1-z)| < \pi$ by the integral

$$F\left({a, b \atop c}; z\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-zt)^a}$$

where $\arg(1-zt) = 0$ for $z \in \mathbb{R}_{<0}$, and for $|z| < 1$ and arbitrary a, b, c by the series

$$F\left({a, b \atop c}; z\right) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, \quad (a)_n := a(a+1)(a+2)\cdots(a+n-1),$$

which implies $F\left({a,b \atop c}; 0\right) = 1$. It satisfies the differential equation

$$x(1-x)y'' + (c - (a+b+1)x)y' - aby = 0, \quad y(x) := {}_2F_1\left({a, b \atop c}; x\right)$$

for $x \notin \{1, \infty\}$.

Lemma 2.8.1. *Let $x \in \mathbb{R}_{\geq 0}$, $\nu \in i\mathbb{R} \cup (-1/2, 1/2)$ and $s \in \mathbb{C}$ with $\mathrm{Re}(s) \geq 1/2$. Then*

$$\left| {}_2F_1\left({\frac{1}{2}+\nu, \frac{1}{2}-\nu \atop s}; -x\right) \right| \leq 1.$$

Proof. Fix ν and s as above, and let

$$F_s(x) = {}_2F_1\left({\frac{1}{2}-\nu, \frac{1}{2}+\nu \atop s}; -x\right)$$

for $x \in \mathbb{R}_{\geq 0}$. Then F_s satisfies the differential equation

$$x(1+x)F_s''(x) + (s+2x)F_s'(x) + \lambda F_s(x) = 0 \quad \text{with } \lambda = \tfrac{1}{4} + r^2 > 0. \tag{2.120}$$

Note that since $\{\overline{\frac{1}{2}+ir}, \overline{\frac{1}{2}-ir}\} = \{\frac{1}{2}+ir, \frac{1}{2}-ir\}$, we have $\overline{F_s} = F_{\bar{s}}$ and $\overline{F_s}' = F_{\bar{s}}'$. Let f be a smooth function on $\mathbb{R}$ and $H = |F_s|^2 + f|F_s'|^2$, so that

$$H' = F_s' F_{\bar{s}} + F_s F_{\bar{s}}' + f'|F_s|^2 + f(F_s'' F_{\bar{s}}' + F_s' F_{\bar{s}}''). \tag{2.121}$$

By the differential equation (2.120), we have

$$H' = (F_s' F_{\bar{s}} + F_s F_{\bar{s}}') \left(1 - f \frac{\lambda}{x(1+x)}\right) + |F_s'|^2 \left(f' - f \frac{s + \bar{s} + 4x}{x(1+x)}\right).$$

Taking $f(x) = x(1+x)/\lambda$ gives

$$H'(x) = \frac{1 - 2\mathrm{Re}(s) - 2x}{\lambda} |F_s'|^2(x),$$

so that $H'(x) \leq 0$ for $\mathrm{Re}(s) \geq 1/2$ and $x \geq 0$. Since $f(0) = 0$ and $f(x) \geq 0$ for $x \geq 0$, we obtain

$$|F_s|^2(x) \leq H(x) \leq H(0) = |F_s|^2(0) = 1,$$

as desired. $\square$

Lemma 2.8.2. *Let $\nu \in i\mathbb{R} \cup (-\frac{1}{2}, \frac{1}{2})$ and $s \in \mathbb{C}$ with $\mathrm{Re}(s) \geq 1$. Then*

$$\left| \frac{\Gamma(s+\nu)\Gamma(s-\nu)}{\Gamma(s+\frac{1}{2})\Gamma(s-\frac{1}{2})} \right| \leq 1.$$

Proof. Recall that Kummer's first formula asserts

$$\frac{\Gamma(s+\nu)\Gamma(s-\nu)}{\Gamma(s+\frac{1}{2})\Gamma(s-\frac{1}{2})} = \lim_{x \to 1^-} F_{\nu,s}(x), \quad F_{\nu,s}(x) := F\left(\begin{matrix} \nu + \frac{1}{2}, \nu - \frac{1}{2} \\ s + \nu \end{matrix} ; x \right). \tag{2.122}$$

Write $\sigma = \mathrm{Re}(s)$ and $u = \mathrm{Re}(\nu)$. Take $H = |F_{\nu,s}|^2 + f|F_{\nu,s}'|^2$ for a smooth function f. The differential equation

$$x(1-x)F_{\nu,s}''(x) + (s + \nu - (2\nu+1)x)F_{\nu,s}'(x) + \lambda F_{\nu,s}(x) = 0,$$

with $\lambda = \frac{1}{4} - \nu^2 > 0$, implies that

$$\begin{aligned} H' = &\left(F_{\nu,s}' F_{\bar{\nu},\bar{s}} + F_{\nu,s} F_{\bar{\nu},\bar{s}}'\right) \left(1 - f \frac{\lambda}{x(1-x)}\right) \\ &+ |F_{\nu,s}'|^2 \left(f' - f \frac{2\sigma + 2u - 2(2u+1)x}{x(1-x)}\right). \end{aligned}$$

Taking $f(x) = x(1-x)/\lambda$ gives

$$H'(x) = \frac{1 - 2\sigma - 2u(1-x) + 2ux}{\lambda} |F'_{s,\nu}|^2(x),$$

so that our hypotheses $u \in (-\frac{1}{2}, \frac{1}{2})$, $\mathrm{Re}(s) \geq 1$ imply $H'(x) \leq 0$ for $0 \leq x < 1$. Since $f(0) = 0$ and $f(x) \geq 0$ for $0 \leq x \leq 1$, we obtain $|F_{s,\nu}|^2(x) \leq H(x) \leq H(0) = |F_{s,\nu}|^2(0) = 1$ for $x \in (0,1)$, and the lemma follows from (2.122). $\square$

Remark 5. The proof of Lemma 2.8.2 shows that the hypothesis $\mathrm{Re}(s) \geq 1$ can be relaxed to $\mathrm{Re}(s) \geq \frac{1}{2} + \mathrm{Re}(\nu)$; we believe that Lemma 2.8.2 holds in the larger range $\mathrm{Re}(s) \geq \frac{1}{2}$, $\nu \in i\mathbb{R} \cup (-\frac{1}{2}, \frac{1}{2})$, but have not proven this. Such refinements are not necessary for our applications in the proof of Lemma 2.4.3.

Remark 6. The bounds asserted by Lemmas 2.8.1 and 2.8.2 are sharp for several extremal cases of the parameters.

Chapter 3

Equidistribution of Cusp Forms In The Level Aspect

3.1 Introduction

3.1.1 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 [25] 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 [52].

Theorem 3.1.1 (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 $Y_0(q)$.*

[1]as spelled out by Luo and Sarnak [42]; we refer to Sarnak [53, 54] and the references in [25] 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 $C_c(X)$ and writing 1 for the constant function.

Kowalski, Michel, and VanderKam [36, 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 3.1.2 (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 3.1.2 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 [70], which is now known by theorem 3.4.1 of this chapter. Recently Koyama [37], following the method of Luo and Sarnak [41], proved the analogue of Conjecture 3.1.2 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 chapter is to establish the squarefree level case of Conjecture 3.1.2. Our result is the first of its kind for nondihedral cusp forms.

Theorem 3.1.3 (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 7. Our extension (theorem 3.4.1) of Watson's formula [70] shows that theorem 3.1.3 would follow from subconvex bounds $L(f \times f \times \phi, 1/2) \ll_\phi q^{1-\delta}$ $(\delta > 0)$ for the central L-values of the triple product L-functions attached to f as above and each Maass cusp form or unitary Eisenstein series ϕ on $Y_0(1)$. Such bounds are known to follow from the generalized Lindelöf hypothesis, which itself follows from the generalized Riemann hypothesis, so one can view theorem 3.1.3 as an unconditionally proven consequence of a central unresolved conjecture.

Remark 8. One cannot relax entirely the restriction of theorem 3.1.3 to newforms, since for instance a cusp form of level 1 may be regarded as an oldform of arbitrary level $q > 1$.

Remark 9. Rudnick [51] showed that theorem 3.1.1 implies that the zeros of newforms of level 1 and weight $k \to \infty$ equidistribute on $Y_0(1)$. At the 2010 Arizona Winter School, Soundararajan asked 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 $q \to \infty$? Following Rudnick, one may show for $\phi \in C_c^\infty(\mathbb{H})$ and $\Phi(z) = \sum_{\gamma \in \Gamma} \phi(\gamma z)$ that

$$\frac{12}{k\psi(q)} \sum_{z \in \Gamma \backslash \mathcal{Z}_1} \frac{\Phi(z)}{\# \operatorname{Stab}_\Gamma(z)} = \int_{\Gamma \backslash \mathbb{H}} \Phi \, dV + \int_{\Gamma \backslash \mathbb{H}} \frac{\pi_{q*}(\log \nu_f)}{k\psi(q)} \Delta \Phi \, dV, \tag{3.1}$$

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 (3.1) 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 $\#\Gamma\backslash\mathcal{Z}_1 = \#\Gamma_0(q)\backslash\mathcal{Z} \sim k\psi(q)/12$ [60, §2], the first term on the right-hand side of (3.1) 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 toward 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 3.1.3 does not seem to preclude the masses ν_f from being very small somewhere within each fiber of the projection $Y_0(q) \to Y_0(1)$; stated another way, the *sum* of the values taken by $y^k|f|^2$ in a fiber of $Y_0(q) \to Y_0(1)$ are controlled (in an average sense as the fiber varies) by theorem 3.1.3, but their *product* could still conceivably be quite small. There are further difficulties in adapting Rudnick's method that we shall not mention here.

Remark 10. Lindenstrauss [40] and Soundararajan [65] 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 3.1.2? An affirmative answer to this question would follow from the generalized Riemann hypothesis (at least for q squarefree, as in remark 7), but appears beyond the reach of our methods because the Ramanujan conjecture is not known for Maass forms (compare with [25, p.2]).

Remark 11. We shall actually 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 $qk \to \infty$. The novelty in our argument concerns only the variation of q, so we encourage the reader to regard k as fixed.

Remark 12. 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 3.3.4, 3.3.13, and 3.4.3. One should be able to generalize Lemmas 3.3.4 and 3.3.13 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 3.4.3, one must compute (or sharply bound) 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.

3.1.2 Plan for the Chapter

Our chapter is organized as follows. In §3.2 we recall some standard properties of our basic objects of study: holomorphic newforms, Maass eigencuspforms, unitary Eisenstein series and incomplete Eisenstein series. In §3.3 we prove the level aspect analogue of Holowinsky's main result [24, Corollary 3], as described above; we emphasize the aspects of his argument that do not immediately generalize to the level aspect and refer to his paper for the details of arguments that do. In §3.4 we extend Watson's formula to cover the additional case that we need. In §3.5 we complete the proof of theorem 3.1.3 using the main results of §3.3 and §3.4. Sections 3.3 and 3.4 are independent of each other, but both depend upon the definitions, notation and facts recalled in §3.2.

3.1.3 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 $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \circlearrowright \mathbb{H}$ acting by fractional linear transformations, 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

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

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

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

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 $f(x,y,z) \ll_{x,y} g(x,y,z) \ll_{x,y} f(x,y,z)$.

3.1.4 Weyl's Criterion

The following standard lemma provides essential motivation for what follows.

Lemma 3.1.4. *Suppose that for each fixed Maass eigencuspform or incomplete Eisenstein series ϕ, we have*

$$\frac{\mu_f(\phi)}{\mu_f(1)} \to \frac{\mu(\phi)}{\mu(1)} \quad \text{as } qk \to \infty$$

for q squarefree and f a holomorphic newform of weight k and level q; the convergence need not be uniform in ϕ. Then theorem 3.1.3 is true.

Proof. The family of probability measures $\phi \mapsto \mu_f(\phi)/\mu_f(1)$ obtained as f varies is equicontinuous for the supremum norm on $C_c(Y_0(1))$, since $|\mu_f(\phi_1)/\mu_f(1) - \mu_f(\phi_2)/\mu_f(1)| \leq \sup|\phi_1 - \phi_2|$ for any bounded functions ϕ_1, ϕ_2 on $Y_0(1)$. Thus theorem 3.1.3 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 set is furnished [29] by the Maass eigencuspforms and incomplete Eisenstein series as defined in §3.2. □

3.1.5 Acknowledgements

We thank Dinakar Ramakrishnan for suggesting this problem and for his very helpful feedback and comments on earlier drafts of this chapter. We thank Abhishek Saha for his careful reading of, and useful comments on, an earlier draft. The problem that we address was raised explicitly by K. Soundararajan at the 2010 Arizona Winter School; we thank him as well as Roman Holowinsky, Henryk Iwaniec, Philippe Michel, and Peter Sarnak for their encouragement. We thank the referees for their numerous helpful suggestions and corrections.

3.2 Background on Automorphic Forms

We collect here some standard properties of classical automorphic forms. We refer to Serre [59], Shimura [60], Iwaniec [28, 29] and Atkin-Lehner [1] for complete definitions and proofs.

3.2.1 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

$$y^{k/2} f(z) = \sum_{n \in \mathbb{N}} \frac{\lambda_f(n)}{\sqrt{n}} \kappa_f(ny) e(nx), \tag{3.2}$$

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 [8, 9] 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 3.1.3 makes sense.

To a newform f one attaches the finite part of the adjoint L-function $L(\mathrm{ad}\, f, s) = \prod_p L_p(\mathrm{ad}\, f, s)$ and its completion $\Lambda(\mathrm{ad}\, f, s) = L_\infty(\mathrm{ad}\, f, s) L(\mathrm{ad}\, f, s) = \prod_v L_v(\mathrm{ad}\, f, s)$, where p traverses the set of primes and v the set of places of $\mathbb{Q}$; the local factors $L_v(\mathrm{ad}\, f, s)$ are as in [70, §3.1.1]. The Rankin-Selberg method [50, 57] and a standard calculation [70, §3.2.1] show that

$$\mu_f(1) := \int_{\Gamma_0(q)\backslash\mathbb{H}} |f|^2(z) y^k \, \frac{dx\, dy}{y^2} = q \frac{\Gamma(k-1)}{(4\pi)^{k-1}} \frac{k-1}{2\pi^2} L(\mathrm{ad}\, f, 1). \tag{3.3}$$

As in the analogous weight aspect [25, p.7], the work of Gelbart-Jacquet [13] (following Shimura [61]) and the theorem of Hoffstein-Lockhart [22, Theorem 0.1] (with appendix by Goldfeld-Hoffstein-Lieman) imply that

$$L(\mathrm{ad}\, f, 1)^{-1} \ll \log(qk). \tag{3.4}$$

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 $d_\sigma :=$

[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.

$[\Gamma_\infty : \Gamma_\infty \cap \sigma\Gamma_0(q)\sigma^{-1}]$ 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 [29, §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, p.138]. Thus by applying [1, Thm 3] to the newform f, we obtain

$$f|_k w_\sigma = \pm f. \tag{3.5}$$

Since f is $\Gamma_0(q)$-invariant, the property (3.5) does not depend upon the choice of coset representative σ.

3.2.2 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 [58] 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

$$\phi(z) = \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{\lambda_\phi(n)}{\sqrt{|n|}} \kappa_{ir}(ny) e(nx) \tag{3.6}$$

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, $\operatorname{sgn}(y) = 1$ or -1 according as y is positive or negative, 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 [29, Theorem 3.2]

$$\sum_{n \leq x} |\lambda_\phi(n)|^2 \ll_\phi x. \tag{3.7}$$

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

3.2.3 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

$$E(s,z) = y^s + M(s)y^{1-s} + \frac{1}{\xi(2s)} \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{\lambda_{s-1/2}(n)}{\sqrt{|n|}} \kappa_{s-1/2}(ny)e(nx), \tag{3.8}$$

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.*

3.2.4 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*

$$E(\Psi, z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \Psi(\operatorname{Im}(\gamma z)). \tag{3.9}$$

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

$$E(\Psi, z) = \int_{(2)} \Psi^\wedge(s) E(s,z) \frac{ds}{2\pi i} = \frac{\Psi^\wedge(1)}{\operatorname{vol}(\Gamma \backslash \mathbb{H})} + \int_{(1/2)} \Psi^\wedge(s) E(s,z) \frac{ds}{2\pi i}. \tag{3.10}$$

Let $\phi = E(\Psi, \cdot)$ be an incomplete Eisenstein series. Note that $\mu(\phi) = \Psi^\wedge(1)$. By comparing (3.10) and (3.8), 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

$$\phi_n(y) = \int_{(1/2)} \frac{\Psi^\wedge(s)}{\xi(2s)} \frac{\lambda_{s-1/2}(n)}{\sqrt{|n|}} \kappa_{s-1/2}(ny) \frac{ds}{2\pi i} \qquad (n \neq 0), \tag{3.11}$$

$$\phi_0(y) = \frac{\mu(\phi)}{\mu(1)} + \int_{(1/2)} \Psi^\wedge(s) \left(y^s + M(s)y^{1-s}\right) \frac{ds}{2\pi i} \qquad (n = 0). \tag{3.12}$$

3.3 Main Estimates

We prove a level aspect analogue of Holowinsky's main bound [24, Corollary 3]. To formulate our result, define for each normalized holomorphic newform f and each real number $x \geq 1$ the quantities

$$M_f(x) = \frac{\prod_{p\leq x}(1 + 2|\lambda_f(p)|/p)}{\log(ex)^2 L(\mathrm{ad}\, f, 1)}, \quad R_f(x) = \frac{x^{-1/2}}{L(\mathrm{ad}\, f, 1)} \int_{\mathbb{R}} \left| \frac{L(\mathrm{ad}\, f, \frac{1}{2} + it)}{(1 + |t|)^{10}} \right| dt. \tag{3.13}$$

In §3.5 we shall refer only to the definitions (3.13) and the statement of the following theorem, not its proof.

Theorem 3.3.1. *Let f be a normalized holomorphic newform of weight k and squarefree level q. If ϕ is a Maass eigencuspform, then*

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

If ϕ is an incomplete Eisenstein series, then

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

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 §3.3.1 we reduce theorem 3.3.1 to a problem of estimating shifted sums (see Definition 3.3.2). In §3.3.2 we apply a refinement of [24, Theorem 2] to bound such shifted sums. In §3.3.3 we complete the proof of theorem 3.3.1.

3.3.1 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 3.3.2. 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 3.3.1 to the problem of bounding such shifted sums. We shall subsequently refer to the statement below of Proposition 3.3.3 but not the details of its proof.

Proposition 3.3.3. *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

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

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 (3.9).

Lemma 3.3.4. *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 (3.10), 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 [24, Proof of Lemma 3.1a] shows without modification that

$$\int_{(1/2)} h^\wedge(s) Y^s \mu_f(E(s,\cdot)\phi)\,\frac{ds}{2\pi i} \ll_\phi Y^{1/2}\mu_f(1); \tag{3.14}$$

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 $h^\wedge(s)Y^s E(s,z)\phi(z) \ll_\phi Y^{1/2}|s|^{-2}$ for s, z as above; the estimate (3.14) 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 §3.2.1, 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 (3.5) 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 3.3.4, 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

$$\sum_{d|q} \int_{y=0}^{\infty} h_Y(dy) \int_{x=0}^{1} \phi(dz)|f|^2(z) y^k \, \frac{dx\,dy}{y^2} = \mathcal{S}_0 + \mathcal{S}_{(0,Y^{1+\varepsilon})} + \mathcal{S}_{\geq Y^{1+\varepsilon}}. \tag{3.15}$$

We treat these contributions in Lemmas 3.3.6, 3.3.7 and 3.3.8, respectively; in doing so we shall repeatedly use the following technical result.

Lemma 3.3.5. *The quantity $\mu_f(E(h_Y,\cdot))$ satisfies the formulas and estimates*

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

where

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

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 3.3.4, the second from the Mellin formula and the unfolding method by a direct computation, the third from the bounds $|\Gamma(k-1/2+it)| \leq \Gamma(k-1/2)| \ll k^{-1/2}\Gamma(k)$, $\zeta(1/2+it) \ll (1+|t|)^{1/4}$ and $|\zeta(1+2it)| \gg 1/\log(1+|t|)$ as in [66, p.7]. 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 intersects $\ll Y$ fundamental domains for $\Gamma\backslash\mathbb{H}$ [29, Lemma 2.10], we have $\mu_f(E(h_Y,\cdot)) \ll Y\mu_f(1)$. □

Lemma 3.3.6 (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*

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

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 (3.12) 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 3.3.5 show that

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

□

Lemma 3.3.7 (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 (3.2) of a newform, the Fourier expansion (3.6) of a Maass cusp form, and the formula (3.11) for the non-constant Fourier coefficients of an Eisenstein series. □

Lemma 3.3.8 (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 3.3.8 follows from Lemma 3.3.5 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 [24, §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 (3.11) 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 $|l| \geq Y^{1+\varepsilon}$. $\square$

Proof of Proposition 3.3.3. By Lemma 3.3.4 and equation (3.15), 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 3.3.3 follows by combining the results of Lemma 3.3.6, Lemma 3.3.8 and Lemma 3.3.7. $\square$

3.3.2 Bounds for Individual Shifted Sums

We bound the individual shifted sums appearing in Definition 3.3.2; in subsequent sections we shall need only our main result, Corollary 3.3.12. We first recall a special case of Holowinsky's bound [24, Theorem 2].

Theorem 3.3.9 (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 3.3.9 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 3.3.10. *With conditions as in the statement of theorem 3.3.9, we have*

$$\sum_{\substack{n \in \mathbb{N} \\ m := n + l \in \mathbb{N} \\ \max(m,n) \leq x}} |\lambda_f(m) \lambda_f(n)| \ll_\varepsilon \frac{x \prod_{p \leq x} (1 + 2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}} \tag{3.16}$$

where all implied constants are absolute.

Proof. In [48, Thm 3.1], we generalized Holowinsky's bound [24, Thm 2] to totally real number fields $\mathbb{F}$. Along the way we proved a pair of results [48, Thm 4.10] and [48, Thm 7.2] either of which imply theorem 3.3.10. For completeness, we shall give the argument here in the special case $\mathbb{F} = \mathbb{Q}$, which borrows heavily from that of Holowinsky; up to (3.20) we essentially recall his argument, and after that introduce our refinement. Let $\lambda(n) = |\lambda_f(n)|$, so that

$$\lambda \text{ is a nonnegative multiplicative function satisfying } \lambda(n) \leq \tau(n). \tag{3.17}$$

We may assume that $1 \leq l \leq x$. Fix $\alpha \in (0, 1/2)$ (to be chosen sufficiently small at the end of the proof) and set

$$y = x^{\alpha}, \quad s = \alpha \log\log(x), \quad z = x^{1/s}.$$

For $x \gg_{\alpha} 1$ we have $10 \leq z \leq y \leq x$, as we shall henceforth assume. For each $n \in \mathbb{N}$, write $m = n + l \in \mathbb{N}$. Define the *z-part* of a positive integer to be the greatest divisor of that integer supported on primes $p \leq z$. There exist unique positive integers a, b, c such that $\gcd(a, b) = 1$ and ac (resp. bc) is the z-part of m (resp. n); such triples a, b, c satisfy

$$p|abc \Rightarrow p \leq z, \quad c|l, \quad \text{and } \gcd(a, b) = 1. \tag{3.18}$$

Write $\mathbb{N} = \sqcup_{a,b,c}\mathbb{N}_{abc}$ for the fibers of $n \mapsto (a, b, c)$. The assumption (3.17) implies $\lambda(m)\lambda(n) \leq 4^s\lambda(ac)\lambda(bc)$, so that

$$\begin{aligned} \sum_{n \in \mathbb{N} \cap [1,x]} \lambda(m)\lambda(n) &= \sum_{a,b,c} \sum_{n \in \mathbb{N}_{abc} \cap [1,x]} \lambda(m)\lambda(n) \\ &\leq 4^s \sum_{a,b,c} \lambda(ac)\lambda(bc) \cdot \#(\mathbb{N}_{abc} \cap [1, x]). \end{aligned}$$

Holowinsky asserts that Rankin's trick implies that the contribution to the above from a, b, c for which $|ac| > y$ or $|bc| > y$ is $\ll_{\alpha,A} x \log(x)^{-A}$ for any A; we spell out an alternate proof of this assertion in [48, Lemma 7.3]. Now, an integer belongs to $\mathbb{N}_{abc}$ only if it satisfies some congruence conditions modulo each prime $p \leq z$ (see [24, p.14], or [48, Lemma 7.3] for a detailed discussion); as in [24] or [48, Corollary 7.8], an application of the large sieve (or Selberg's sieve) shows that if $|ac| \leq y$, $|bc| \leq y$ and $x \gg y^2$, then[4]

$$\#(\mathbb{N}_{abc} \cap [1, x]) \ll \frac{x + (yz)^2}{\log(z)^2} \frac{l}{c^2\phi(abc^{-1}l)}. \tag{3.19}$$

Since $(yz)^2 \ll_{\alpha} x$, $\log(z)^2 \asymp_{\alpha} \log\log(x)^{-2}\log(x)^2$, $4^s \ll_{\varepsilon} \log(x)^{\varepsilon}$ (for $\alpha \ll_{\varepsilon} 1$), and $\phi(abc^{-1}l) \geq$

[4]This bound is slightly poorer than that obtained by Holowinsky because we have been more precise in our calculation of the residue classes sieved out by prime divisors of $c^{-1}l$; the discrepancy here does not matter in the end.

$\phi(c^{-1}l)\phi(a)\phi(b)$, we see that theorem 3.3.10 follows from the bound[5]

$$\sum_{\substack{c|l \\ p|c \Rightarrow p \leq z}} \frac{1}{c} \frac{l/c}{\phi(l/c)} \sum_{\substack{|ac| \leq y \\ p|ab \Rightarrow p \leq z}} \sum_{|bc| \leq y} \frac{\lambda(ac)\lambda(bc)}{\phi(a)\phi(b)} \ll \log\log(x)^{O(1)} \prod_{p \leq z} \left(1 + \frac{2\lambda(p)}{p}\right), \tag{3.20}$$

which we now establish. Note first that

$$\sum_{\substack{|ac| \leq y \\ p|ab \Rightarrow p \leq z}} \sum_{|bc| \leq y} \frac{\lambda(ac)\lambda(bc)}{\phi(a)\phi(b)} \leq \left(\prod_{p \leq z} \sum_{k \geq 0} \frac{\lambda(p^{k+v_p(c)})}{\phi(p^k)} \right)^2. \tag{3.21}$$

Using that $\lambda(p^k) \leq k+1$ and $p \geq 2$ and summing some geometric series as in [48, Lemma 7.4] gives

$$\sum_{k \geq 0} \frac{\lambda(p^{k+\nu})}{\phi(p^k)} \leq \nu + 1 + \sum_{k \geq 1} \frac{\nu + k + 1}{p^{k-1}(p-1)} \leq 3\nu + 3$$

for each $\nu \geq 1$, while for $\nu = 0$

$$\begin{aligned} \sum_{k \geq 0} \frac{\lambda(p^k)}{\phi(p^k)} &= \left(1 + \frac{\lambda(p)}{p}\right) \left(1 + \frac{\lambda(p)\left(\frac{1}{\phi(p)} - \frac{1}{p}\right) + \sum_{k \geq 2} \frac{\lambda(p^k)}{\phi(p^k)}}{1 + \frac{\lambda(p)}{p}}\right) \\ &\leq \left(1 + \frac{\lambda(p)}{p}\right)\left(1 + \frac{20}{p}\right). \end{aligned}$$

Thus the LHS of (3.20) is bounded by $\zeta(2)^{40}\psi(l)\prod_{p \leq z}(1 + \lambda(p)p^{-1})^2$, where ψ is the multiplicative function

$$\psi(l) = \sum_{c|l} \frac{1}{c} \frac{l/c}{\phi(l/c)} \prod_{p^\nu || c} (3\nu + 3)^2. \tag{3.22}$$

By direct calculation and the inequality $p \geq 2$, we have

$$\psi(p^a) = \frac{1}{1 - p^{-1}} + \frac{9}{p^a} \left((a+1)^2 + \frac{1}{1 - p^{-1}} \sum_{i=1}^{a-1} \frac{(i+1)^2}{p^i} \right) \leq 1 + Cp^{-1}$$

for some constant $C \leq 10^6$, so that $\psi(l) \leq \prod_{p|l}(1 + Cp^{-1}) \ll \log\log(x)^C$ for $1 \leq l \leq x$. This estimate for $\psi(l)$ establishes the claimed bound (3.20). $\square$

Remark 13. A bound of the form (3.16) but with an *unspecified* dependence on the parameter l may be derived from the work of Nair [45]. We have attempted to quantify this dependence

[5] It is here that Holowinsky gives up the factor $\tau(l)$.

by working through the details of Nair's arguments, and have shown that they imply

$$\sum_{\substack{n\in\mathbb{N}\\ m:=n+l\in\mathbb{N}\\ \max(m,n)\leq x}} |\lambda_f(m)\lambda_f(n)| \ll_\varepsilon \tau_m(l)\frac{x\prod_{p\leq x}(1+2|\lambda_f(p)|/p)}{\log(ex)^{2-\varepsilon}} \tag{3.23}$$

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 (3.16) subject to the additional constraint $4l^2 \leq x$ appears in the recent book of Iwaniec-Friendlander [10, Thm 15.6], which was released after we completed the work of this chapter. 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 3.3.2 that the sums $S_s(l,x)$ involve a certain integral $I_s(l,n,x)$.

Lemma 3.3.11. *For each positive integer A, the integral $I_s(l,n,x)$ satisfies the upper bound*

$$I_s(l,n,x) \ll_A \frac{\Gamma(k-1)}{(4\pi)^{k-1}}\sqrt{mn}\cdot\max\left(1,\frac{\max(m,n)}{xk}\right)^{-A}$$

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

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

Here we have used the arithmetic mean-geometric mean inequality, the well-known bound [72, Ch 7, Misc. Ex 44]

$$\frac{\Gamma(w+k-1)}{\Gamma(k-1)} \ll_A (k-1)^A(1+k^{-1}(1+|w|^2)) \ll k^A(1+|w|^2)$$

for $\mathrm{Re}(w) = A$, and the rapid decay of $h^\wedge$. The case $A = 0$ gives $I_s(l,n,x) \ll_k (4\pi)^{-k+1}\Gamma(k-1)\sqrt{mn}$, which combined with the case that A is a positive integer yields the assertion of the lemma. □

Remark 14. See [48, Lem 4.3] and [48, Cor 4.4] for a fairly sharp refinement of Lemma 3.3.11.

Corollary 3.3.12. *The shifted sums $S_s(l,x)$ satisfy the upper bound*

$$S_s(l,x) \ll_\varepsilon \frac{\Gamma(k-1)}{(4\pi)^{k-1}} \frac{xk}{\log(xk)^{2-\varepsilon}} \prod_{p \leq xk} \left(1 + \frac{2|\lambda_f(p)|}{p}\right) \tag{3.24}$$

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

Proof. Let us set $X = xk$ and temporarily denote by $T_f(x,l,\varepsilon)$ the right-hand side of (3.24) without the factor $(4\pi)^{-k+1}\Gamma(k-1)$. By Definition 3.3.2 and Lemma 3.3.11, we need only show that

$$\sum_{\substack{n \in \mathbb{N} \\ m := n+l \in \mathbb{N}}} |\lambda_f(m)\lambda_f(n)| \cdot \max\left(1, \frac{\max(m,n)}{X}\right)^{-A} \ll_\varepsilon T_f(x,l,\varepsilon) \tag{3.25}$$

for some positive integer A. Take $A = 2$. We may assume that $X = xk \geq 10$. By theorem 3.3.10 and the Deligne bound $|\lambda_f(p)| \leq 2$, the left hand side of (3.25) is

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

The inner sum converges and is bounded uniformly in X, so we obtain the desired estimate (3.25). □

3.3.3 Bounds for Sums of Shifted Sums

We complete the proof of theorem 3.3.1 by bounding the sums of shifted sums that arose in Proposition 3.3.3.

Lemma 3.3.13. *For each $\varepsilon \in (0,1)$ and each squarefree number q, we have*

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

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 $i = 1, \ldots, r$. Define $\beta(x) = x/\log(e^e xk)^{2-\varepsilon}$; we have chosen this particular definition so that β is increasing on $\mathbb{R}_{\geq 1}$ and $\beta(x) \asymp x/\log(xk)^{2-\varepsilon}$ 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 $a = (a_1, \ldots, a_r) \in \{0,1\}^r$ we have

$$\frac{\beta(q_1^{a_1} \cdots q_r^{a_r})}{\beta(q_1 \cdots q_r)} \leq \frac{\beta(p_1^{a_1} q_2^{a_2} \cdots q_r^{a_r})}{\beta(p_1 q_2 \cdots q_r)} \leq \frac{\beta(p_1^{a_1} p_2^{a_2} q_3^{a_3} \cdots q_r^{a_r})}{\beta(p_1 p_2 q_3 \cdots q_r)} \leq \cdots \leq \frac{\beta(p_1^{a_1} \cdots p_r^{a_r})}{\beta(p_1 \cdots p_r)}. \tag{3.26}$$

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

$$\begin{aligned} p_{r-r_0+1} \cdots p_r &= \exp\bigl(r\log(r) - (r - r_0)\log(r - r_0) + o(r\log(r))\bigr) \\ &= \exp\left(r_0 \log(r) + (r - r_0)\log\left(\frac{r}{r - r_0}\right) + o(r\log(r))\right) \\ &= \exp\left(r_0 \log(r)(1 + o(1))\right) \\ &\ll (p_1 \cdots p_r)^{1/9}, \end{aligned} \tag{3.27}$$

and

$$\log(p_1 \cdots p_{r_0}) = r_0 \log(r_0)(1 + o(1)) \asymp r\log(r)(1 + o(1)) = \log(p_1 \cdots p_r). \tag{3.28}$$

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 (3.27) we have

$$\sum_{a \in \Omega_0} \frac{\beta(p_1^{a_1} \cdots p_r^{a_r})}{\beta(p_1 \cdots p_r)} \leq 2^r \frac{\beta(p_{r-r_0+1} \cdots p_r)}{\beta(p_1 \cdots p_r)} \ll 2^r (p_1 \cdots p_r)^{-7/8} \leq \sqrt[8]{2}. \tag{3.29}$$

If $a \in \Omega_1$, then (3.28) 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

$$\sum_{a \in \Omega_1} \frac{\beta(p_1^{a_1} \cdots p_r^{a_r})}{\beta(p_1 \cdots p_r)} \ll \sum_{d | p_1 \cdots p_r} \frac{1}{d} \leq (1 + o(1)) e^{\gamma} \log\log(p_1 \cdots p_r) \ll \log\log(e^e q). \tag{3.30}$$

Since $\beta(x) \asymp x/\log(ex)^{2-\varepsilon}$ for $x \in \mathbb{R}_{\geq 1}$, it follows from (3.26), (3.29), and (3.30) that

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

which establishes the lemma. $\square$

Corollary 3.3.14. *Let $Y \geq 1$ with $Y \leq c_1 \log(qk)^{c_2}$ for some $c_1, c_2 \geq 1$. Then our sum of* *shifted sums satisfies the estimate*

$$\sum_{d|q} S_s(dl, dY) \ll_{\varepsilon, c_1, c_2} \frac{\Gamma(k-1)}{(4\pi)^{k-1}} \frac{qkY}{\log(qk)^{2-\varepsilon}} \prod_{p \le qk} \left(1 + \frac{2|\lambda_f(p)|}{p}\right),$$

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

Proof. By Corollary 3.3.12, we have

$$\sum_{d|q} S_s(dl, dY) \ll_\varepsilon \frac{\Gamma(k-1)}{(4\pi)^{k-1}} Y \left(\prod_{p \le qkY} \left(1 + 2\frac{|\lambda_f(p)|}{p}\right)\right) \sum_{d|q} \frac{dk}{\log(dk)^{2-\varepsilon}}. \tag{3.31}$$

By the Deligne bound $|\lambda_f(p)| \leq 2$, the part of the product in (3.31) taken over $qk < p \leq qkY$ is $\ll \log(eY)^4 \ll_{c_1,c_2} \log\log(e^e qk)^4$. The claim now follows from Lemma 3.3.13. □

Lemma 3.3.15. *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 (3.7) for λ_ϕ and the uniform bound $|\lambda_{it}(l)| \leq \tau(l)$ for λ_{it}. □

Proof of theorem 3.3.1. Suppose that ϕ is a normalized Maass eigencuspform of eigenvalue $\frac{1}{4} + r^2$. By Proposition 3.3.3, we have

$$\frac{\mu_f(\phi)}{\mu_f(1)} = \frac{1}{Y\mu_f(1)} \sum_{0<|l|<Y^{1+\varepsilon}} \frac{\lambda_\phi(l)}{\sqrt{|l|}} \sum_{d|q} S_{ir}(dl, dY) + O_{\phi,\varepsilon}(Y^{-1/2}). \tag{3.32}$$

Recall from (3.3) that

$$\mu_f(1) \asymp q\frac{\Gamma(k-1)}{(4\pi)^{k-1}} L(\mathrm{ad}\, f, 1)$$

and recall the definition (3.13) of $M_f(qk)$. We shall ultimately choose $Y \ll \log(qk)^{O(1)}$, so Corollary 3.3.14 gives the bound

$$\frac{1}{Y\mu_f(1)} \sum_{d|q} S_{ir}(dl, dY) \ll_\varepsilon \log(qk)^\varepsilon M_f(qk). \tag{3.33}$$

By (3.33) and Lemma 3.3.15 applied to (3.32), we find that

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

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

Suppose now that $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series. Proposition 3.3.3, Corollary 3.3.14 and Lemma 3.3.15 show, as in the cuspidal case, that

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

The same choice of Y as above completes the proof. $\square$

3.4 An Extension of Watson's Formula

Watson [70], building on earlier work of Garrett [11], Piatetski-Shapiro and Rallis [49], Harris and Kudla [19], and Gross and Kudla [18], 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 §3.5 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 3.4.1. *Let ϕ be a Maass eigencuspform of level 1 and f a holomorphic newform of squarefree level q, as in §3.2. Then*

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

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 [70, §3].

Remark 15. For simplicity, we have stated theorem 3.4.1 only in the special case that we need it, but our calculations (Lemma 3.4.3) lead to a 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 an even or odd Maass eigencuspform. If $k_1 + k_2 + k_3 \neq 0$ or some prime p divides exactly one of the q_j, then it is straightforward to see that $\int \psi_1\psi_2\psi_3 = 0$. Otherwise $k_1+k_2+k_3 = 0$ and each prime divides the q_j either $0, 2$ or 3 times, so one can read off from Watson [70, Theorem 3], Ichino [26] and Lemma 3.4.3 the identity

$$\frac{\left|\int_X \psi_1\psi_2\psi_3\right|^2}{\prod \int_X |\psi_j|^2} = \frac{1}{8}\frac{\Lambda(\frac{1}{2}, \psi_1 \times \psi_2 \times \psi_3)}{\prod \Lambda(1, \operatorname{ad} \psi_j)} \prod_v c_v \tag{3.34}$$

where $X = \varprojlim \Gamma_0(q)\backslash\mathbb{H}$ with $\mathrm{vol}(X) := \mathrm{vol}(\Gamma_0(1)\backslash\mathbb{H}) = \pi/3$, c_∞ is $Q_\infty \in \{0, 1, 2\}$ from [70, Theorem 3], $c_p = 1$ if p divides none of the q_j, $c_p = p^{-1}$ if p divides exactly two of the q_j, and $c_p = p^{-1}(1 + p^{-1})(1 + \varepsilon_p)$ if p divides all of the q_j with $-\varepsilon_p$ the product of the Atkin-Lehner eigenvalues for the ψ_j at p as in [70, Theorem 3].

Watson proved his formula only for three forms of the same squarefree level because Gross and Kudla [18] evaluated the p-adic zeta integrals of Harris and Kudla [19] 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 [26] 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 [27, §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 [4] 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. In what follows, v denotes a place of $\mathbb{Q}$ and p a 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 $\bar{f}_v = \left(\begin{smallmatrix} -1 & \\ & 1 \end{smallmatrix}\right) \cdot f_v$ 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

$$\tilde{I}_v = \left(\frac{\zeta_v(2)^3}{\zeta_v(2)} \frac{L_v(\frac{1}{2}, \phi \times f \times f)}{L_v(1, \mathrm{ad}\, \phi) L_v(1, \mathrm{ad}\, f)^2} \right)^{-1} I_v. \tag{3.35}$$

Theorem 3.4.2 (Ichino)**.** *We have $\tilde{I}_v = 1$ for all but finitely many places v, and*

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

Proof. See [26, 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 [12]) and the comparison (see for instance Vignéras [69, §III.2]) between the Poincaré measure on the upper half-plane and the Tamagawa measure on $G_{\mathbb{A}}$. $\square$

We know by work of Harris and Kudla [19], Gross and Kudla [18], Watson [70], Ichino [27], and Ichino and Ikeda [27] 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, with which we deduce theorem 3.4.1 from theorem 3.4.2.

Lemma 3.4.3. *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 $\beta = \beta_\phi(p)$. 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 [5, Theorem 4.6.6]

$$\Phi_{\phi,p}(a(p^m)) = \frac{1}{1+p^{-1}} p^{-m/2} \left[\alpha^m \frac{1 - p^{-1}\frac{\beta}{\alpha}}{1 - \frac{\beta}{\alpha}} + \beta^m \frac{1 - p^{-1}\frac{\alpha}{\beta}}{1 - \frac{\alpha}{\beta}} \right]. \tag{3.36}$$

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 [15, §7] (see also [27, §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 $\tilde{W} = \langle w_1, w_2, \omega \rangle$, which is the semidirect product of W_a by the group of order 2 generated by ω, 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

$$\Phi_{f,p}(\omega^j w)^2 = p^{-2\lambda(w)}. \tag{3.37}$$

In the general case that $\pi_{f,p}$ is a possibly nontrivial unramified quadratic twist of Steinberg, the formula (3.37) for the *squared* matrix coefficient still holds.

Proof of Lemma 3.4.3. Having recalled the formulas above, we see that

$$\begin{aligned} I_p &= \int_{G_p} \Phi_{\phi,p}(g) \Phi_{f,p}(g)^2 \, dg = \sum_{w \in \tilde{W}} \mathrm{vol}(I_p w I_p) \Phi_{\phi,p}(w) p^{-2\lambda(w)} \qquad (3.38) \\ &= (p+1)^{-1} \sum_{w \in \tilde{W}} \Phi_{\phi,p}(w) p^{-\lambda(w)}, \end{aligned}$$

where $\Phi_{\phi,p}$ is given by (3.36). The evaluation of the Poincaré series

$$\sum_{w \in \tilde{W}} t^{\lambda(w)} = 2 \frac{1+t}{1-t}, \tag{3.39}$$

where t is an indeterminate, is asserted and used in [27, §7], but we need a finer 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

$$\sum_{w\in\tilde{W}} x^{\mu(w)}t^{\lambda(w)} = \frac{(1+x)(1+t)}{1-xt}. \tag{3.40}$$

Note that we recover (3.39) upon taking $x = 1$. To prove (3.40), observe that since $\omega w_1 = w_2\omega$ and $\omega^2 = 1$, every element w of $\tilde{W}$ is of the form $u_{abn} = \omega^a(w_1w_2)^n w_1^b$ or $v_{abn} = \omega^a(w_2w_1)^n w_2^b$ 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

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

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

$$v_{00n} = \begin{pmatrix} p^{-n} & \\ & p^n \end{pmatrix}, \quad v_{01n} = \begin{pmatrix} & p^{-n-1} \\ p^{n+1} & \end{pmatrix},$$

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

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 (3.40) follows upon summing the geometric series. We now combine (3.36), (3.38) and (3.40), noting that the series converge because $|\alpha| < p^{1/2}$ and $|\beta| < p^{1/2}$; the contributions to the formula (3.38) for I_p of the two terms in the formula (3.36) 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 (3.35) of $\tilde{I}_p$. The local L-factors are given by (see [70, §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. □

3.5 Proof of Theorem 3.1.3

We combine theorem 3.3.1 and theorem 3.4.1 with Soundararajan's weak subconvex bounds [66] to complete the proof of theorem 3.1.3. 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 $qk \to \infty$, thereby fulfilling the criterion of Lemma 3.1.4, by combining the complementary estimates for $D_f(\phi)$ provided below by Proposition 3.5.2 and Proposition 3.5.3.

Lemma 3.5.1. *The quantities $M_f(x)$ and $R_f(x)$ (3.13) appearing in the statement of theorem 3.3.1 satisfy the estimates*

$$M_f(qk) \ll_\varepsilon \log(qk)^{1/6+\varepsilon} L(\operatorname{ad} f, 1)^{1/2}, \quad R_f(qk) \ll_\varepsilon \frac{\log(qk)^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)} \ll \log(qk)^\varepsilon.$$

Proof. The bound for $M_f(qk)$ follows from the proof of [25, Lemma 3] with "k" replaced by "qk," noting that $\lambda_f(p)^2 \leq 1 + \lambda_f(p^2)$ for *all* primes p. The bound for $R_f(qk)$ follows from the arguments of [66, Example 1], [25, Lemma 1] with "k" replaced by "qk" and the lower bound

(3.4) for $L(\mathrm{ad}\, f, 1)$. $\square$

Proposition 3.5.2. *We have* $D_f(\phi) \ll_{\phi,\varepsilon} \log(qk)^{1/12+\varepsilon} L(\mathrm{ad}\, f, 1)^{1/4}$.

Proof. Follows immediately from theorem 3.3.1 and Lemma 3.5.1. $\square$

Proposition 3.5.3. *We have* $D_f(\phi) \ll_{\phi,\varepsilon} \log(qk)^{-\delta+\varepsilon} L(\mathrm{ad}\, f, 1)^{-1}$*, 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 $\asymp (qk)^4$, so theorem 3.4.1 and the arguments of Soundararajan [66, Example 2] with "k" replaced by "qk" show that

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

If $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series, then the unfolding method as in Lemma 3.3.5 and the bound for $R_f(q)$ given by Lemma 3.5.1 show that

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

$\square$

Proof of theorem 3.1.3. By Propositions 3.5.2 and 3.5.3, there exists $\delta \in \{1/2, 1\}$ such that

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

it follows by the argument of [25, §3] with "k" replaced by "qk" that $D_f(\phi) \to 0$ as $qk \to \infty$. $\square$

Bibliography

[1] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160. MR MR0268123 (42 #3022)

[2] Don Blasius, *Hilbert modular forms and the Ramanujan conjecture*, Noncommutative Geometry and Number Theory, Aspects Math., E37, Vieweg, Wiesbaden, 2006, pp. 35–56. MR MR2327298 (2008g:11077)

[3] Valentin Blomer and Gergely Harcos, *Twisted L-functions over number fields and Hilbert's eleventh problem*, Geom. Funct. Anal. **20** (2010), no. 1, 1–52. MR 2647133

[4] S. Böcherer and R. Schulze-Pillot, *On the central critical value of the triple product L-function*, Number Theory (Paris, 1993–1994), London Math. Soc. Lecture Note Ser., vol. 235, Cambridge University Press, Cambridge, 1996, pp. 1–46. MR MR1628792 (99j:11051)

[5] Daniel Bump, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics, vol. 55, Cambridge University Press, Cambridge, 1997. MR MR1431508 (97k:11080)

[6] Y. Colin de Verdière, *Ergodicité et fonctions propres du laplacien*, Bony-Sjöstrand-Meyer seminar, 1984–1985, École Polytech., Palaiseau, 1985, pp. Exp. No. 13, 8. MR MR819779 (87h:58219)

[7] Harold Davenport, *Multiplicative number theory*, second ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 1980, Revised by Hugh L. Montgomery. MR MR606931 (82m:10001)

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

[9] ———, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. **43** (1974), 273–307. MR MR0340258 (49 #5013)

[10] John Friedlander and Henryk Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, vol. 57, American Mathematical Society, Providence, RI, 2010. MR 2647984

[11] Paul B. Garrett, *Decomposition of Eisenstein series: Rankin triple products*, Ann. of Math. (2) **125** (1987), no. 2, 209–235. MR MR881269 (88m:11033)

[12] Stephen Gelbart, *Automorphic forms on adèle groups*, Princeton University Press, Princeton, N.J., 1975, Annals of Mathematics Studies, No. 83. MR MR0379375 (52 #280)

[13] Stephen Gelbart and Hervé Jacquet, *A relation between automorphic representations of* GL(2) *and* GL(3), Ann. Sci. École Norm. Sup. (4) **11** (1978), no. 4, 471–542. MR MR533066 (81e:10025)

[14] ______, *Forms of* GL(2) *from the analytic point of view*, Automorphic Forms, Representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 213–251. MR MR546600 (81e:10024)

[15] Roger Godement and Hervé Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Mathematics, Vol. 260, Springer-Verlag, Berlin, 1972. MR MR0342495 (49 #7241)

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, seventh ed., Elsevier/Academic Press, Amsterdam, 2007, Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX). MR MR2360010 (2008g:00005)

[17] George Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 43, Springer-Verlag, Berlin, 2001. MR MR1836967 (2002i:11092)

[18] Benedict H. Gross and Stephen S. Kudla, *Heights and the central critical values of triple product L-functions*, Compositio Math. **81** (1992), no. 2, 143–209. MR MR1145805 (93g:11047)

[19] Michael Harris and Stephen S. Kudla, *The central critical value of a triple product L-function*, Ann. of Math. (2) **133** (1991), no. 3, 605–672. MR MR1109355 (93a:11043)

[20] Adolf Hildebrand and Gérald Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux **5** (1993), no. 2, 411–484. MR MR1265913 (95d:11116)

[21] Jürgen G. Hinz, *Methoden des grossen Siebes in algebraischen Zahlkörpern*, Manuscripta Math. **57** (1987), no. 2, 181–194. MR 871630 (88d:11113)

[22] Jeffrey Hoffstein and Paul Lockhart, *Coefficients of Maass forms and the Siegel zero*, Ann. of Math. (2) **140** (1994), no. 1, 161–181, With an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman. MR MR1289494 (95m:11048)

[23] Roman Holowinsky, *A sieve method for shifted convolution sums*, Duke Math. J. **146** (2009), no. 3, 401–448. MR 2484279 (2010b:11127)

[24] ______, *Sieving for mass equidistribution*, Ann. of Math. (2) **172** (2010), no. 2, 1499–1516. MR 2680498

[25] Roman Holowinsky and Kannan Soundararajan, *Mass equidistribution for Hecke eigenforms*, Ann. of Math. (2) **172** (2010), no. 2, 1517–1528. MR 2680499

[26] Atsushi Ichino, *Trilinear forms and the central values of triple product L-functions*, Duke Math. J. **145** (2008), no. 2, 281–307. MR MR2449948 (2009i:11066)

[27] Atsushi Ichino and Tamutsu Ikeda, *On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture*, Geom. Funct. Anal. **19** (2010), no. 5, 1378–1425. MR MR2585578

[28] Henryk Iwaniec, *Topics in classical automorphic forms*, Graduate Studies in Mathematics, vol. 17, American Mathematical Society, Providence, RI, 1997. MR MR1474964 (98e:11051)

[29] ______, *Spectral methods of automorphic forms*, second ed., Graduate Studies in Mathematics, vol. 53, American Mathematical Society, Providence, RI, 2002. MR MR1942691 (2003k:11085)

[30] Henryk Iwaniec, *Notes on the quantum unique ergodicity for holomorphic cusp forms*, 2010.

[31] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004. MR MR2061214 (2005h:11005)

[32] Henryk Iwaniec and Peter Sarnak, *Perspectives on the analytic theory of L-functions*, Geom. Funct. Anal. (2000), no. Special Volume, Part II, 705–741, GAFA 2000 (Tel Aviv, 1999). MR MR1826269 (2002b:11117)

[33] Hervé Jacquet, *Automorphic forms on* GL(2)*. Part II*, Lecture Notes in Mathematics, Vol. 278, Springer-Verlag, Berlin, 1972. MR MR0562503 (58 #27778)

[34] Hervé Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Lecture Notes in Mathematics, Vol. 114, Springer-Verlag, Berlin, 1970. MR MR0401654 (53 #5481)

[35] E. Kowalski, *The large sieve and its applications*, Cambridge Tracts in Mathematics, vol. 175, Cambridge University Press, Cambridge, 2008, Arithmetic geometry, random walks and discrete groups. MR MR2426239 (2009f:11123)

[36] E. Kowalski, P. Michel, and J. VanderKam, *Rankin-Selberg L-functions in the level aspect*, Duke Math. J. **114** (2002), no. 1, 123–191. MR MR1915038 (2004c:11070)

[37] Shin-ya Koyama, *Equidistribution of Eisenstein series in the level aspect*, Comm. Math. Phys. **289** (2009), no. 3, 1131–1150. MR MR2511663

[38] Uwe Krause, *Abschätzungen für die Funktion* $\Psi_K(x,y)$ *in algebraischen Zahlkörpern*, Manuscripta Math. **69** (1990), no. 3, 319–331. MR MR1078363 (91i:11165)

[39] J.-P. Labesse and R. P. Langlands, *L-indistinguishability for* SL(2), Canad. J. Math. **31** (1979), no. 4, 726–785. MR MR540902 (81b:22017)

[40] Elon Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) **163** (2006), no. 1, 165–219. MR MR2195133 (2007b:11072)

[41] Wenzhi Luo and Peter Sarnak, *Quantum ergodicity of eigenfunctions on* $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 207–237. MR MR1361757 (97f:11037)

[42] ———, *Mass equidistribution for Hecke eigenforms*, Comm. Pure Appl. Math. **56** (2003), no. 7, 874–891, Dedicated to the memory of Jürgen K. Moser. MR MR1990480 (2004e:11038)

[43] S. Marshall, *Mass Equidistribution for Automorphic Forms of Cohomological Type on GL_2*, ArXiv e-prints (2010).

[44] H. L. Montgomery, *A note on the large sieve*, J. London Math. Soc. **43** (1968), 93–98. MR 0224585 (37 #184)

[45] Mohan Nair, *Multiplicative functions of polynomial values in short intervals*, Acta Arith. **62** (1992), no. 3, 257–269. MR MR1197420 (94b:11093)

[46] Mohan Nair and Gérald Tenenbaum, *Short sums of certain arithmetic functions*, Acta Math. **180** (1998), no. 1, 119–144. MR 1618321 (99j:11110)

[47] Paul Nelson, *Equidistribution of cusp forms in the level aspect*, Duke Math. J. (to appear).

[48] ———, *Mass equidistribution of Hilbert modular eigenforms*, Ramanujan J. (to appear).

[49] I. Piatetski-Shapiro and Stephen Rallis, *Rankin triple L functions*, Compositio Math. **64** (1987), no. 1, 31–115. MR MR911357 (89k:11037)

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

[51] Zeév Rudnick, *On the asymptotic distribution of zeros of modular forms*, Int. Math. Res. Not. (2005), no. 34, 2059–2074. MR MR2181743 (2006k:11099)

[52] Zeév Rudnick and Peter Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. **161** (1994), no. 1, 195–213. MR MR1266075 (95m:11052)

[53] Peter Sarnak, *Arithmetic quantum chaos*, The Schur lectures (1992) (Tel Aviv), Israel Math. Conf. Proc., vol. 8, Bar-Ilan Univ., Ramat Gan, 1995, pp. 183–236. MR MR1321639 (96d:11059)

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

[55] Werner Schaal, *On the large sieve method in algebraic number fields*, J. Number Theory **2** (1970), 249–270. MR 0272745 (42 #7626)

[56] A. I. Schnirelman, *Ergodic properties of eigenfunctions*, Uspehi Mat. Nauk **29** (1974), no. 6(180), 181–182. MR MR0402834 (53 #6648)

[57] Atle Selberg, *Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist*, Arch. Math. Naturvid. **43** (1940), 47–50. MR MR0002626 (2,88a)

[58] ———, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., Providence, R.I., 1965, pp. 1–15. MR MR0182610 (32 #93)

[59] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7. MR MR0344216 (49 #8956)

[60] 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. MR MR0314766 (47 #3318)

[61] ———, *On the holomorphy of certain Dirichlet series*, Proc. London Math. Soc. (3) **31** (1975), no. 1, 79–98. MR MR0382176 (52 #3064)

[62] ———, *The special values of the zeta functions associated with Hilbert modular forms*, Duke Math. J. **45** (1978), no. 3, 637–679. MR MR507462 (80a:10043)

[63] Lior Silberman and Akshay Venkatesh, *On quantum unique ergodicity for locally symmetric spaces*, Geom. Funct. Anal. **17** (2007), no. 3, 960–998. MR MR2346281 (2009a:81072)

[64] K. Soundararajan, *Arizona winter school lecture notes on quantum unique ergodicity and number theory*, `http://math.arizona.edu/~swc/aws/10/2010SoundararajanNotes.pdf`, 2010.

[65] Kannan Soundararajan, *Quantum unique ergodicity for* $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, Ann. of Math. (2) **172** (2010), no. 2, 1529–1538. MR 2680500

[66] ______, *Weak subconvexity for central values of L-functions*, Ann. of Math. (2) **172** (2010), no. 2, 1469–1498. MR 2680497

[67] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, second ed., The Clarendon Press Oxford University Press, New York, 1986, Edited and with a preface by D. R. Heath-Brown. MR MR882550 (88c:11049)

[68] Akshay Venkatesh, *Sparse equidistribution problems, period bounds and subconvexity*, Ann. of Math. (2) **172** (2010), no. 2, 989–1094. MR 2680486

[69] Marie-France Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics, vol. 800, Springer, Berlin, 1980. MR MR580949 (82i:12016)

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

[71] André Weil, *Séries de Dirichlet et fonctions automorphes*, Séminaire Bourbaki, Vol. 10, Soc. Math. France, Paris, 1995, pp. Exp. No. 346, 547–552. MR MR1610485

[72] E. T. Whittaker and G. N. Watson, *A course of modern analysis. An introduction to the general theory of infinite processes and of analytic functions: with an account of the principal transcendental functions*, Fourth edition. Reprinted, Cambridge University Press, New York, 1962. MR MR0178117 (31 #2375)

[73] Steven Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55** (1987), no. 4, 919–941. MR MR916129 (89d:58129)