MASS EQUIDISTRIBUTION OF HILBERT MODULAR EIGENFORMS

PAUL D. NELSON

ABSTRACT. Let $\mathbb{F}$ be a totally real number field, and let f traverse a sequence of non-dihedral 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.

CONTENTS

1. Introduction

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 paper 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 §1.2, comes from its connection to quantum chaos by analogy with the *quantum unique ergodicity* conjecture of Rudnick and Sarnak [35] 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 [14] 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 1.1. *Let $f : \mathrm{PGL}_2(\mathbb{A}) \to \mathbb{C}$ traverse a sequence of non-dihedral 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*

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

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 1.1 asserts that $|f|^2 \, d\mu$ *converges weakly* to $d\mu$. Theorem 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 $\max(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty$ on the union of the remaining connected components of Y remains true, but to simplify the exposition we shall consider only non-dihedral forms in this paper.

We proved a slightly weaker form of Theorem 1.1 in September 2009 and learned soon thereafter from Sarnak's lecture notes [37] that overlapping results had been obtained earlier that year in the 2009/2010 Princeton Ph.D thesis of his student Marshall [30]. We hope that our own arguments differ sufficiently to be of interest.

The case $\mathbb{F} = \mathbb{Q}$ of Theorem 1.1 is the celebrated theorem of Holowinsky-Soundararajan [14], who established a quantitative rate of convergence in the limit (1) for a "spanning set" of functions ϕ (see §3). Marshall [30] 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 $\min(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty$ with $\min(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \geq (k_1 \cdots k_{[\mathbb{F}:\mathbb{Q}]})^\eta$ for some fixed $\eta > 0$; he remarks that the former assumption concerning the class number is not essential. 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 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 5.5.

An important ingredient in Holowinsky's contribution to proof of Theorem 1.1 when $\mathbb{F} = \mathbb{Q}$ is his bound

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

---

[1] those induced from idele class characters on unramified totally imaginary quadratic extensions of $\mathbb{F}$; see §2.8.1

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 §3.1). A generalization of (2) to number fields features in Marshall's work mentioned above. We independently generalize (2) to number fields 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) and its generalizations (see Theorem 4.10 and Theorem 7.2). 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 [34].

1.2. **Motivation.** The study of the limiting behavior of the masses of Hilbert modular eigencuspforms is natural and interesting from several perspectives of which we highlight two. 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

$$(3) \qquad\qquad (\Delta + \lambda)\phi = 0$$

of the Laplacian Δ on a compact Riemannian manifold M for which the geodesic flow is chaotic (see [36]). Here the geodesic flow on M is regarded as the Hamiltonian flow of a chaotic classical mechanical system, the Laplacian $\Delta \circlearrowleft L^2(M)$ 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 (3) as $\lambda \to \infty$ is akin to considering the semi-classical 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 [42, 3, 51] 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.[2] In the particular case that M is negatively curved, the *quantum unique ergodicity* (QUE) conjecture of Rudnick and Sarnak [35] 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 [37, 28, 27, 44, 41, 14]. 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 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 representative model cases for the more general problem (see [36]).

The variant of QUE that we consider for holomorphic Hilbert modular eigencuspforms f of increasing weight is in the spirit of the original conjectures and was spelled out explicitly for the modular curve ($\mathbb{F} = \mathbb{Q}$) by Luo and Sarnak [29]; it is important here that f is taken to be an eigenform (of the Hecke algebra), since for instance the powers of a fixed form have weight tending to infinity but do not have equidistributed mass.

A second motivation for our considerations arises from their connection to central problems in the analytic theory of L-functions. Watson [48] 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 §3.2). 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

---

[2]in a more precise sense than we describe here; see the introduction to [35]

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.

1.3. **Overview of proof.** Recall that we consider non-dihedral 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 1.1 follows as soon as one can establish (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 §2.8. To highlight the essential difficulties let us suppose in this section that ϕ is a Maass eigencuspform. Then $\int \phi = 0$, so to establish (1) we must show that

$$(4) \qquad \frac{\int \phi |f|^2}{\int |f|^2} \to 0 \quad \text{as } \max(k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]}) \to \infty,$$

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 (4) by a remarkable synthesis of their independent efforts [15, 43], which we now recall briefly, saving a more detailed discussion for §3 and referring to the lucid expositions of [14, 37, 45] for further motivation and details. Watson's formula [48] and work of Gelbart-Jacquet [7] and Hoffstein-Lockhart-Goldfeld-Lieman [13] imply (see [14, Lem 2]) that

$$(5) \qquad \frac{\int \phi |f|^2}{\int |f|^2} \approx_\phi \frac{|L(\phi \times \mathrm{ad}\, f, \frac{1}{2})|^{1/2}}{k^{1/2}} \exp\left(-\sum_{p \le k} \frac{1}{p} \lambda(p^2)\right),$$

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 (4) provided that

$$(6) \qquad \frac{\sum_{p \le k} \frac{1}{p} \lambda(p^2)}{\sum_{p \le k} \frac{1}{p}} \ge -1/2 + \delta + o_{k \to \infty}(1) \quad \text{for some fixed } \delta > 0.$$

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 [19])

$$(7) \qquad \frac{\int \phi |f|^2}{|f|^2} \ll_{\phi, \varepsilon} \log(k)^\varepsilon \exp\left(-\sum_{p \le k} \frac{1}{p}(|\lambda(p)| - 1)^2\right),$$

which implies (4) provided that

$$(8) \qquad \frac{\sum_{p \le k} \frac{1}{p}(|\lambda(p)| - 1)^2}{\sum_{p \le k} \frac{1}{p}} \ge \delta + o_{k \to \infty}(1) \quad \text{for some fixed } \delta > 0.$$

In summary, Soundararajan succeeds unless typically $\lambda(p^2) \lessgtr -1/2$, 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

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

so in all cases one 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 §3, the only part of the $\mathbb{F} = \mathbb{Q}$ argument that does not generalize transparently is Holowinsky's proof of (7). His argument amounts to

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

$$(9) \qquad X^{-1} \sum_{n \in \mathbb{Z} \cap [1, X]}^{\mathrm{smooth}} \lambda(n)\lambda(n + l),$$

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

(2) bounding the shifted sums (9); a reformulation [19] of the bound that Holowinsky obtains is

$$(10) \qquad X^{-1} \sum_{n, n+l \in \mathbb{Z} \cap [1, X]} |\lambda(n)\lambda(n + l)| \ll \tau(l) \log(k)^{\varepsilon} \prod_{p \leq k} \left(1 + \frac{2(|\lambda(p)| - 1)}{p} \right),$$

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 (9) 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 [29, 14, 37, 45] and especially [16] 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 (9) 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 (10) fail to produce an estimate of the quality (7) 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 4.5) to the precise asymptotic expansion

$$(11) \qquad \frac{\int \phi |f|^2}{\int |f|^2} \sim \frac{(Yk)^{-1}}{L(\mathrm{ad}\, f, 1)} \sum_{\substack{m = n + l \\ \max(m, n) \asymp Yk}} \sum \frac{\lambda_{\phi}(l)}{\sqrt{|l|}} \lambda_f(m) \lambda_f(n) \kappa_{\phi, \infty} \left(\frac{k-1}{4\pi} \left| \log \frac{m}{n} \right| \right),$$

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 4.3 and Corollary 4.4; see also Remark 4.7) what amounts to the overwhelming decay of the Bessel factor $\kappa_{\phi, \infty}(\cdots)$ in the higher-dimensional generalization of (11) 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 §6). In this way we reduce to bounding shifted sums of the form (9) 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 (10) and (7) 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 upon these last few paragraphs in successively greater detail in §3 and §4.

1.4. **Plan for the paper.** In §2 we introduce notation that will allow us to speak meaningfully about automorphic forms over totally real fields. In §3 we review the work of Holowinsky and Soundararajan over $\mathbb{F} = \mathbb{Q}$ and reduce the proof of our main result Theorem 1.1 to that of a generalization (Theorem 3.1) of Holowinsky's bound (7). The heart of our paper is §4, in which we prove Theorem 3.1 assuming some independent technical results that we relegate to §5, §6 and §7.

1.5. **Acknowledgements.** We thank Dinakar Ramakrishnan for suggesting this problem and for his very helpful feedback and comments on earlier drafts of this paper. We also thank Fokko van de Bult for a conversation that lead to a strengthening and simplification of the proof of Lemma 6.1.

2. Preliminaries

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 $e_{\mathbb{F}_\infty}(x) = e^{2\pi i \, \mathrm{Tr}(x)}$), $\mathbb{F}_{\infty+}^*$ 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 $\mathrm{div}\, \alpha \in I_{\mathbb{F}}$ 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 $\mathrm{div}\, d_{\mathbb{F}} = \mathfrak{d}$ and $\mathrm{div}\, z_j = \mathfrak{z}_j$ for $j = 1, \ldots, h(\mathbb{F})$. Then we have natural identifications

$$(12) \qquad \mathbb{A}^* = \sqcup_{j=1}^{h(\mathbb{F})} \mathbb{F}^* (\mathbb{F}_{\infty+}^* \times z_j^{-1} \hat{\mathfrak{o}}^*), \quad \mathbb{F}^* \backslash \mathbb{A}^* / \hat{\mathfrak{o}}^* = \sqcup_{j=1}^{h(\mathbb{F})} \left((\mathbb{F}_{\infty+}^* / \mathfrak{o}_+^*) \times z_j^{-1} \right).$$

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

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 upon 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 upon ε and A without mention.

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

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

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

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

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 $k = (k_1, \ldots, k_d) \in (2\mathbb{Z}_{\geq 1})^d$ 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 $i \in \{1, \ldots, d\}$; in particular, $x > \mathbf{0}$ signifies that $x_i > 0$ for all i, i.e., that x is totally positive.

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

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

and the accompanying notation

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

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

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

and let $K = K_\infty \times K_{\mathrm{fin}}$. 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 [40] 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 $g = n(x)a(y)kz$.

2.5. 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

$$(13) \qquad \int_{Z(\mathbb{A})B(\mathbb{Q})G(\mathbb{A})} \phi = \int_{x\in\mathbb{F}\backslash\mathbb{A}} \int_{y\in\mathbb{F}^*\backslash\mathbb{A}^*} \int_{k\in K} \phi(n(x)a(y)k)\, dx \, \frac{d^\times y}{|y|_\mathbb{A}} \, dk$$

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 $d^\times t = t^{-1}\, dt$.

2.6. 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

$$(14) \qquad y \mapsto \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} \mathrm{sgn}(y_j)^{\varepsilon_j}|y_j|^{ir_j} \quad \text{if } y = (y_1, \ldots, y_{[\mathbb{F}:\mathbb{Q}]}) \in (\mathbb{R}^{[\mathbb{F}:\mathbb{Q}]})^* = \mathbb{F}_\infty^*$$

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 (14); 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

$$(15) \qquad \Psi(y) = \int_{\mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)(c)} \Psi^\wedge(\chi)\chi(y)\,\frac{d\chi}{2\pi i}$$

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 §2.5), the forward transform is given explicitly by

$$(16) \qquad \Psi^\wedge(\chi) = \frac{1}{\mathrm{vol}(C_\mathbb{F}^1)} \int_{C_\mathbb{F}} \Psi(y)\chi^{-1}(y)\,d^\times y.$$

The analytic conductor [21] of an unramified idele class character $\chi \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)$ having archimedean component (14) is defined to be

$$(17) \qquad C(\chi) = \prod_{i=1}^{[\mathbb{F}:\mathbb{Q}]} (3 + |r_j|);$$

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

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

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.

2.7. **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

$$(18) \qquad \phi(n(x)a(y)) = \phi_0(y) + \sum_{n \in \mathbb{F}^*} \kappa_\phi(ny)e_\mathbb{F}(nx)$$

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 [49]).

We say that the Fourier expansion (18) of ϕ is *factorizable* if for each $y \times z \in \mathbb{F}_\infty^* \times \mathbb{A}_f^* = \mathbb{A}^*$ we have

$$(19) \qquad \kappa_\phi(y \times z) = \kappa_{\phi,\infty}(y)\frac{\lambda_\phi(\mathrm{div}\, z)}{\mathrm{N}(\mathrm{div}\, z)^{1/2}},$$

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 $\kappa_{\phi,\infty_j} : \mathbb{R}^* \to \mathbb{C}$.

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

2.8.1. *Holomorphic eigencuspforms.* By a *holomorphic eigencuspform* $f : \mathbf{X} \to \mathbb{C}$ of weight $k = (k_1, \ldots, k_{[\mathbb{F}:\mathbb{Q}]})$ (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 [40]; 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 (18) with $f_0 \equiv 0$ and

$$(20) \qquad \kappa_{f,\infty_j}(y) = \begin{cases} y^{k_j/2} e^{-2\pi y} & \text{for } y > 0, \\ 0 & \text{for } y < 0 \end{cases}$$

for each infinite place ∞_j of $\mathbb{F}$. The "Ramanujan bound" for f [1] asserts[3] that $|\lambda_f(\mathfrak{a})| \le \tau(\mathfrak{a})$ 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 $|\lambda_f(\mathfrak{p})| \le 2$ 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 [48, §3.1.1]. It is known [39, 6] 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 *non-dihedral* 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 quadratic extension of $\mathbb{F}$ [6, 26]. 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.

2.8.2. *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 (18) with $\phi_0 \equiv 0$ and

$$(21) \qquad \kappa_{\phi,\infty_j}(y) = 2|y|^{1/2} K_{ir_j}(2\pi|y|) \operatorname{sgn}(y)^{\varepsilon_j}$$

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 $\lambda_\phi(\mathfrak{a}) \le \tau(\mathfrak{a}) \operatorname{N}(\mathfrak{a})^{1/2}$. The "Rankin-Selberg bound," also known as the "Ramanujan bound on average," asserts that

$$(22) \qquad \sum_{\operatorname{N}(\mathfrak{a}) \le x} |\lambda_\phi(\mathfrak{a})|^2 \ll_\phi x$$

and follows as in [18, §8.2] from the analytic properties of the Rankin-Selberg L-series attached to $\phi \times \phi$ [23].

---

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

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*

$$E(\chi, g) = \sum_{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F})} \chi(y(\gamma g)) \tag{23}$$

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 [7]). 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 (18) with

$$\phi_0(y) = \Delta_{\mathbb{F}}^{-1} \chi(d_{\mathbb{F}})^{-2} \xi_{\mathbb{F}}(\chi^2) \chi(y) + \Delta_{\mathbb{F}}^{-1/2} \xi_{\mathbb{F}}(\chi^2 |.|^{-1}) \chi^{-1}(y)|y|, \tag{24}$$

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

as in §2.7, where for $\chi \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)$ with χ_∞ given by (14), we set

$$\kappa_{\chi_{\infty_j}}(y) = 2|y|^{1/2} K_{ir_j}(2\pi|y|) \mathrm{sgn}(y)^{\varepsilon_j}, \quad \lambda_\chi(\mathfrak{p}^k) = \sum_{i=0}^{k} \chi(\mathfrak{p})^i \chi^{-1}(\mathfrak{p})^{k-i}; \tag{25}$$

for a convenient tabulation of such Fourier expansions of Eisenstein series see [2].

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 $|\lambda_{\chi|.|^{-1/2}}(\mathfrak{a})| \leq \tau(\mathfrak{a})$.

2.8.4. *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

$$E(\Psi, g) = \sum_{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F})} \Psi(y(\gamma g)) \tag{26}$$

with $y(\gamma g)$ as in §2.8.3. Write $\phi = E(\Psi, \cdot)$. We have $\Psi^\wedge(|.|) \mathrm{res}_{s=1} E(|.|^s, \cdot) = \mu(\phi)/\mu(1)$ (see §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 [7] and [18, §7.3]), we obtain

$$E(\Psi, g) = \frac{\mu(\phi)}{\mu(1)} + \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0) \chi_0(\det g)$$
$$+ \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi) E(\chi, g) \frac{d\chi}{2\pi i} \tag{27}$$

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

$$\phi_0(y) = \frac{\mu(\phi)}{\mu(1)} + \sum_{1 \neq \chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0) \chi_0(y) + O_\phi(|y|^{1/2}), \tag{28}$$

$$\kappa_\phi(y \times z) = \int_{\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)(0)} \frac{\Psi^\wedge(|.|^{1/2}\chi)}{\xi_{\mathbb{F}}(|\cdot|\chi^2)\chi(d_{\mathbb{F}})^{-2}} \kappa_{\chi,\infty}(y) \frac{\lambda_\chi(\mathrm{div}\, z)}{\mathrm{N}(\mathrm{div}\, z)^{1/2}} \frac{d\chi}{2\pi i}. \tag{29}$$

2.9. **Masses.** Recall the measure μ defined on the space $\mathbf{X} = Z(\mathbb{A})G(\mathbb{F})\backslash G(\mathbb{A})$ in §2.5. For $\phi \in L^1(\mathbf{X}, \mu)$ let $\mu(\phi) = \int_{\mathbf{X}} \phi \, d\mu$. To our varying non-dihedral 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 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 §2.8.2 and §2.8.4.

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

$$(30) \qquad \mu_f(1) = \frac{\mathbf{\Gamma}(k)}{c_1(\mathbb{F})(4\pi\mathbf{1})^{k-\mathbf{1}}} L(\mathrm{ad}\, f, 1), \quad c_1(\mathbb{F}) := \frac{(4\pi^2)^{[\mathbb{F}:\mathbb{Q}]}}{2\Delta_{\mathbb{F}}^{3/2}}.$$

We sketch the standard calculation. Recall the measure normalization (13) and the choice of compact subgroup K (§2.4) on which we base our definition (§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 [48, §3.2.1]. Since the Fourier expansion (24) 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 [48, §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 (30).

3. Brief review of Holowinsky-Soundararajan

In this section we summarize the Holowinsky-Soundararajan [14] proof of Theorem 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 [15], and
(2) the independent arguments of Soundararajan [43],
(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 §2.3 that $k^{\mathbf{1}} := k_1 \ldots k_{[\mathbb{F}:\mathbb{Q}]}$, thus when $\mathbb{F} = \mathbb{Q}$ we have $k = (k_1)$ and $k^{\mathbf{1}} = k_1$.

3.1. **Holowinsky's independent arguments.** We begin by simultaneously recalling Holowinsky's main result [15, Cor 3] and stating our generalization thereof. Define for each holomorphic eigencuspform f and each real number $x \geq 2$ the quantities

$$(31) \qquad M_f(x) = \frac{\log(x)^{-2}}{L(\mathrm{ad}\, f, 1)} \prod_{\mathrm{N}(\mathfrak{p}) \leq x} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})} \right),$$

$$(32) \qquad R_f(x) = \frac{x^{-1/2}}{L(\mathrm{ad}\, f, 1)} \sum_{\chi_0 \in \mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} \int_{(1/2)} \left| \frac{L(\mathrm{ad}\, f, \chi_0 |.|^s)}{C(\chi_0 |.|^s)^{10}} \right| |ds|.$$

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

Theorem 3.1. *Let f be a non-dihedral 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 3.1 in §4 by combining the independent results of §5, §6 and §7; doing so is our main task in this paper. Holowinsky [15, Cor 3] established the case $\mathbb{F} = \mathbb{Q}$ of Theorem 3.1, in which the "non-dihedral" 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 [15, Theorem 1] for any $Y \geq 1$ and $\varepsilon > 0$ the asymptotic formula

$$(33) \qquad \frac{\int \phi |f|^2}{\int |f|^2} = cY^{-1} \sum_{0 < |l| < Y^{1+\varepsilon}} S_l(Y) + O_{\phi,\varepsilon}(Y^{-1/2})$$

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

$$(34) \qquad \frac{S_l(Y)}{Y} \ll_{\phi,\varepsilon} \frac{|\phi_l(a(Y^{-1}))|}{L(\mathrm{ad}\, f, 1)} \left[\frac{1}{Yk} \sum_{\substack{n \in \mathbb{N} \\ m := n + l \in \mathbb{N}}} |\lambda_f(m) \lambda_f(n)| h\left(\frac{Y\left(\frac{k-1}{4\pi}\right)}{\frac{m+n}{2}} \right) + \frac{(Yk)^\varepsilon}{k} \right].$$

He then proves [15, 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

$$(35) \qquad \sum_{n \leq x} |\lambda_f(m) \lambda_f(n)| \ll \tau(l) \frac{x}{\log(x)^{2-\varepsilon}} \prod_{p \leq X} \left(1 + \frac{2|\lambda_f(p)|}{p} \right).$$

From this he deduces the cuspidal case of Theorem 3.1 for $\mathbb{F} = \mathbb{Q}$. We generalize and refine (33), (34) and (35) in §5, §6 and §7, respectively; among other refinements, we show that (a generalization to totally real number fields of) the bound (35) holds without the factor $\tau(l)$. The main complication is the manner in which these ingredients fit together to yield Theorem 3.1 when $\mathbb{F} \neq \mathbb{Q}$; this is the crux of our argument, which we present in §4. Specifically, recall from §1.3 that for a totally real number field $\mathbb{F}$ of degree $d = [\mathbb{F} : \mathbb{Q}]$, our naïve generalization of (33) and (34) 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 (34), we reduce to the more tractable problem of showing that a sum of roughly $x \log(x)^\varepsilon$ terms is small relative to x.

3.2. **Soundararajan's independent arguments.** Let ϕ be a Maass eigencuspform, and suppose that $\mathbb{F} = \mathbb{Q}$. Watson's formula [48, Theorem 3] asserts that

$$(36) \qquad \left| \frac{\mu_f(\phi)}{\mu_f(1)} \right|^2 = c(\mathbb{F}, \phi) \frac{\Lambda(\phi \times f \times f, \frac{1}{2})}{\Lambda(\operatorname{ad} f, 1)^2}$$

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 [48, §3.1.1]. The identity (36) with $c(\mathbb{F}, \phi) \neq 0$ holds for totally real $\mathbb{F}$ by Ichino's general triple product formula [17] together with Watson's calculations of the local zeta integrals of Harris-Kudla [10] at the real places. When $\mathbb{F} = \mathbb{Q}$, Soundararajan [43, Ex 2] proves that

$$(37) \qquad L(\phi \times f \times f, \tfrac{1}{2}) \ll_{\phi, \varepsilon} \frac{k^{\mathbf{1}}}{\log(k^{\mathbf{1}})^{1-\varepsilon}}.$$

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

$$(38) \qquad \frac{\int \phi |f|^2}{\int |f|^2} \ll_{\phi, \varepsilon} \frac{\log(k^{\mathbf{1}})^{-1/2+\varepsilon}}{L(\operatorname{ad} f, 1)}.$$

Now let $\phi = E(\chi, \cdot)$ be the unitary Eisenstein series associated as in §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 [43, p7] shows by the unfolding method, Stirling's formula and his weak subconvex bounds for $L(\operatorname{ad} f, \chi)$ [43, Ex 1] – the last of which makes use of the known Ramanujan bound for f – that

$$(39) \qquad \frac{\mu_f(\phi)}{\mu_f(1)} \ll_\varepsilon C(\chi)^2 \frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}$$

and [43, p2]

$$(40) \qquad |L(\operatorname{ad} f, \chi)| \ll_\varepsilon \frac{(k^{\mathbf{1}})^{1/2} C(\chi)^{3/4}}{\log(k^{\mathbf{1}})^{1-\varepsilon}}.$$

By the modularity of $L(\operatorname{ad} f, \chi)$ as the L-function of an automorphic form on $GL(3)$ [6], its Rankin-Selberg theory, and the lower bound

$$(41) \qquad L(\operatorname{ad} f, 1) \gg \log(k^{\mathbf{1}})^{-1}$$

due to Hoffstein-Lockhart-Goldfeld-Hoffstein-Lieman [13] (which is available for general $\mathbb{F}$, see [2, §2.9]), Soundararajan deduces [14, Lem 1] in his joint paper with Holowinsky that

$$(42) \qquad R_f(k^{\mathbf{1}}) \ll_\varepsilon \frac{\log(k^{\mathbf{1}})^\varepsilon}{\log(k^{\mathbf{1}})L(\operatorname{ad} f, 1)} \ll \log(k^{\mathbf{1}})^\varepsilon.$$

The same argument establishes (39), (40), (42) for general totally real number fields $\mathbb{F}$.

3.3. The Holowinsky-Soundararajan synthesis.
In their joint work [14], Holowinsky and Soundararajan show [15, Lem 3] for $\mathbb{F} = \mathbb{Q}$ that

$$(43) \qquad M_k(f) \ll \log(k^{\mathbf{1}})^{1/6} \log\log(k^{\mathbf{1}})^{9/2} L(\operatorname{ad} f, 1)^{1/2},$$

and their proof applies for general $\mathbb{F}$. Subsituting the bound (43) into Theorem 3.1 and combining with Soundararajan's estimate (38) yields for each Maass eigencuspform ϕ that

$$(44) \qquad \frac{\mu_f(\phi)}{\mu_f(1)} \ll_{\phi,\varepsilon} \min\left(\frac{\log(k^{\mathbf{1}})^{-1/2+\varepsilon}}{L(\operatorname{ad} f, 1)}, \log(k^{\mathbf{1}})^{1/12+\varepsilon} L(\operatorname{ad} f, 1)^{1/4}\right).$$

It follows as in [14, 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 3.1.

Holowinsky and Soundararajan show [14, p10] that Soundararajan's bound (39) 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

$$(45) \qquad \left|\frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)}\right| \ll_{\phi,\varepsilon} \frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\operatorname{ad} f, 1)}.$$

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 §2.6)

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

and the meromorphic nature of $E(\chi, \cdot)$ (see §2.8.3 or [7]), we have

$$(46) \qquad \begin{aligned} \mu_f(\phi) = & \sum_{\chi_0 \in \mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)[2]} \Psi^\wedge(\chi_0) \operatorname{res}_{s=1} \mu_f(E(\chi_0|.|^s, \cdot)) \\ & + \int_{\mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi) \mu_f(E(\chi, \cdot)) \frac{d\chi}{2\pi i}, \end{aligned}$$

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 §2.9, 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 non-dihedral in the sense of §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 $\phi \in L^2(\mathbf{X}, \mu)$ [7, §4]. Thus for f non-dihedral, we obtain

$$(47) \qquad \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} = \int_{\mathfrak{X}(C_\mathbb{F}/\hat{\mathfrak{o}}^*)(1/2)} \Psi^\wedge(\chi) \frac{\mu_f(E(\chi, \cdot))}{\mu_f(1)} \frac{d\chi}{2\pi i}.$$

Soundararajan's bound (39) for unitary Eisenstein series shows that the right hand side of (47) is

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

where in the final step we invoked the rapid decay of $\Psi^\wedge$ (see §2.6). Thus we obtain the estimate (45) for non-dihedral forms over a totally real field.

By combining Holowinsky's Theorem 3.1 with Soundararajan's (42) and (45), Holowinsky and Soundararajan obtain, for $\mathbb{F} = \mathbb{Q}$ and $\phi = E(\Psi, \cdot)$, the bound

$$(48) \qquad \left| \frac{\mu_f(\phi)}{\mu_f(1)} - \frac{\mu(\phi)}{\mu(1)} \right| \ll_{\phi,\varepsilon} \min\left(\frac{\log(k^{\mathbf{1}})^{-1+\varepsilon}}{L(\mathrm{ad}\, f, 1)}, \log(k^{\mathbf{1}})^{1/12+\varepsilon} L(\mathrm{ad}\, f, 1)^{1/4} \right),$$

which is $o(1)$ (or even $\ll \log(k^{\mathbf{1}})^{-2/15+\varepsilon}$) by examination (see [14, Proof of Thm 1]). The same estimate follows in the totally real case as soon as one has established Theorem 3.1.

4. The key arguments in our generalization

We saw in §3 that our main result Theorem 1.1 follows from the generalization of Holowinsky's work asserted by Theorem 3.1. We now describe the key arguments that reduce our proof of Theorem 3.1 to several technical results that we shall prove in the remaining sections of this paper; 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 3.1 claims to bound $\mu_f(\phi)/\mu_f(1) - \mu(\phi)/\mu(1)$, for f a non-dihedral 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 $R_f(k^{\mathbf{1}})$ (31)–(32).

Definition 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 $h^\wedge(1) \mathrm{res}_{s=1} E(s, \cdot) = 1$. Recall from §2.1 that we have fixed representatives $\mathfrak{z}_j = \mathrm{div}\, z_j$ for the narrow class group of $\mathbb{F}$; here $j \in \{1, \ldots, h(\mathbb{F})\}$ 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*

$$(49) \qquad S_\chi(x) = \sum_{j=1}^{h(\mathbb{F})} \sum_{\substack{l \in \mathfrak{o}_+^* \backslash \mathfrak{z}_j \\ 0 \neq |l^{\mathbf{1}}| < x^{1+\varepsilon}}} \frac{\lambda_\chi(\mathfrak{z}_j^{-1} l)}{\mathrm{N}(\mathfrak{z}_j^{-1} l)^{1/2}} S_{\chi_\infty}(\mathfrak{z}_j, l, x),$$

where

$$(50) \qquad S_{\chi_\infty}(\mathfrak{z}, l, x) = \sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^* \\ m := n + l \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^*}} \frac{\lambda_f(\mathfrak{z}^{-1} m)}{\mathrm{N}(\mathfrak{z}^{-1} m)^{1/2}} \frac{\lambda_f(\mathfrak{z}^{-1} n)}{\mathrm{N}(\mathfrak{z}^{-1} n)^{1/2}} \frac{I_{\chi_\infty}(l, n, \mathrm{N}(\mathfrak{z})x)}{\mathrm{N}(\mathfrak{z})}$$

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

$$(51) \qquad I_{\chi_\infty}(l, n, x) = \frac{(4\pi\mathbf{1})^{k-1}}{\Gamma(k-1)} \int_{\mathbb{F}_{\infty+}^*} h(xy^{\mathbf{1}}) \kappa_{\chi,\infty}(ly) \kappa_{f,\infty}(my) \kappa_{f,\infty}(ny) \frac{d^\times y}{y^{\mathbf{1}}}.$$

If ϕ is a Maass eigencuspform of eigenvalue $(\frac{1}{4} + r_1^2, \ldots, \frac{1}{4} + r_{[\mathbb{F}:\mathbb{Q}]}^2)$ and parity $(\varepsilon_1, \ldots, \varepsilon_{[\mathbb{F}:\mathbb{Q}]})$, 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 $[y \mapsto \prod \mathrm{sgn}(y_j)^{\varepsilon_j} |y_j|^{ir_j}] \in \mathfrak{X}(\mathbb{F}_\infty^*)$ as in (14).

Proposition 4.2. *Let f be as in the statement of Theorem 3.1 and let $Y \geq 1$. If ϕ is a Maass eigencuspform, then*

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

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

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

The constant $c_1(\mathbb{F})$ is as in the formula (30).

Proof. See §5. The proof is a straightforward and naïve generalization of Holowinsky's arguments in the $\mathbb{F} = \mathbb{Q}$ case. $\qquad\square$

Proposition 4.2 shows that Theorem 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 $\mathfrak{z} = \mathfrak{z}_j$ ($j \in \{1, \ldots, \mathbb{F}\}$), each nonzero shift $l \in \mathfrak{z} \cap \mathbb{F}^*$, and each character $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$; recall from Definition 4.1 that

$$(52) \qquad\qquad\qquad S_{\phi_\infty}(\mathfrak{z}, l, x) = S_{\chi_\infty}(\mathfrak{z}, l, x)$$

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 (14) 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

$$(53) \qquad\qquad I_{\chi_\infty}(l, n, x) = \int_{(c)} h^\wedge(s) x^s \left(\prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]} J_{ir_j}(l_j, n_j, s)\right) \frac{ds}{2\pi i},$$

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

$$(54) \qquad\qquad |J_{ir_j}(l_j, n_j, s)| \leq \frac{\Gamma(k_j - 1 + \sigma)}{\Gamma(k_j - 1)} \frac{\sqrt{m_j n_j}}{\left(4\pi\left(\frac{m_j + n_j}{2}\right)\right)^\sigma} \left(\frac{\sqrt{m_j n_j}}{\left(\frac{m_j + n_j}{2}\right)}\right)^{k_j - 1},$$

where $s = \sigma + it$. However, (54) would *not suffice* for our purposes, as we shall explain after proving the following refinement.

Lemma 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*

$$(55) \qquad\qquad |J_{ir_j}(l_j, n_j, s)| \leq \frac{\Gamma(k_j - 1 + \sigma)}{\Gamma(k_j - 1)} \frac{\sqrt{m_j n_j}}{(4\pi \max(m_j, n_j))^\sigma} \left(\frac{\min(m_j, n_j)}{\max(m_j, n_j)}\right)^{\frac{k_j - 1}{2}}.$$

Proof. By the integral formula [8, 6.621.3] and the transformation formula [8, 9.131] in Gradshteyn-Ryzhik, we have explicitly

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

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 §6, the factors on the second and third lines of (56) are each bounded in absolute value by 1, so the claim follows from the basic inequality $|\Gamma(k_j - 1 + s)| \le \Gamma(k_j - 1 + \sigma)$. $\qquad \square$

Corollary 4.4. *Let $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$ be of the form* (14) *with each $ir_j \in i\mathbb{R} \cup (-\frac{1}{2}, \frac{1}{2})$. Then*

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

Proof. Substitute (55) into (53), taking $c \in \{0, A\}$ and invoking the well known estimate $\Gamma(k_j - 1 + \sigma)/\Gamma(k_j - 1) \ll_\sigma k_j^\sigma$ [50, Ch 7, Misc. Ex 44]. $\qquad \square$

Remark 4.5. With more effort (e.g., by studying the asymptotics of the expression (56)) 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 $x \in \mathbb{F}_{\infty+}^* \cong (\mathbb{R}_+^*)^{[\mathbb{F}:\mathbb{Q}]}$)

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

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

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

This refinement is not necessary for our purposes, so we omit the proof; the simpler upper bound given by Corollary 4.4 suffices because we do not exploit cancellation in the shifted sums, and has the advantage of being completely uniform in χ_∞.

Corollary 4.6. *Let* $\chi_\infty \in \mathfrak{X}(\mathbb{F}_\infty^*)$ *satisfy the hypotheses of Corollary 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*

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

Proof. Substitute Corollary 4.4 into Definition 4.1. $\qquad\square$

Remark 4.7. When $\mathbb{F} = \mathbb{Q}$, Holowinsky applies what amounts to the trivial bound (54), which gives something like (58) upon replacing

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

He then bounds the factor on the RHS of (59) by 1. Now, bounding either of the factors in (59) is harmless when $\mathbb{F} = \mathbb{Q}$: if f has weight k, then in the sum (58) we typically have $m, n \asymp kY$, so for $|l| = O(1)$ both factors in (59) 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 (58) 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 (59) 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 4.4. In fact, the key to our whole argument is that the factor on the LHS of (59) is very small if *any component* of $\mathbf{max}(m, n)$ is not too large, as we quantify in Lemma 4.9.

Definition 4.8. 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

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

Lemma 4.9. *Let* $\chi \in \mathfrak{X}(\mathbb{F}_{\infty+}^*)$ *satisfy the hypotheses of Corollary 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*

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

Proof. We work with the bound asserted by Corollary 4.6. Partition those m, n in (58) 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 (61). It remains to consider those m, n for which

$$
(62) \qquad \max(m, n)_i \leq T_i/U.
$$

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

$$\left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)}\right)^{\frac{k-1}{2}} \leq \exp\left(-\sum_{i=1}^{d} \frac{k_i - 1}{2} \frac{|l_i|}{\max(m_i, n_i)}\right),$$

so the assumption (62) implies

(63)
$$\left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)}\right)^{\frac{k-1}{2}} \leq \exp\left(-\frac{|l_i|U}{3Y^{1/d}}\right).$$

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 (63) we obtain

(64)
$$\left(\frac{\mathbf{min}(m,n)}{\mathbf{max}(m,n)}\right)^{\frac{k-1}{2}} \ll_A X^{-A}.$$

By the trivial "Hecke" bound $\lambda_f(\mathfrak{a}) \ll \mathrm{N}(\mathfrak{a})^{1/2+\varepsilon}$, the contribution to (58) of n satisfying (62) is

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

(65)
$$\ll X^{-A'} \sum_{\substack{n \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^* \\ m := n+l \in \mathfrak{z} \cap \mathbb{F}_{\infty+}^*}} (m^{\mathbf{1}} n^{\mathbf{1}})^{1/2+\varepsilon} \min\left(1, \frac{X}{\mathbf{max}(m,n)^{\mathbf{1}}}\right)^A$$

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 (65) is $\ll_{A''} X^{-A''}$ for any positive constant A'', as desired. $\square$

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 4.9 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 [15, Thm 2].

Theorem 4.10. *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 4.8. 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*

(66)
$$\Sigma_\lambda(\mathfrak{z}, l, T, U) \ll_{\mathfrak{z},\varepsilon} \frac{\log(eU)^{d-1}X}{\log(eX)^{2-\varepsilon}} \prod_{\mathrm{N}(\mathfrak{p}) \leq X} \left(1 + \frac{2|\lambda(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right).$$

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

Proof. See §7. $\square$

Remark 4.11. Holowinsky [15, Thm 2] established a slightly weaker form of the case $d = 1$ of Theorem 4.10 by an application of the large sieve; in his inequality (2) an additional factor of $\tau(l)$ appears on the RHS. We prove Theorem 4.10 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 [15, Thm 2] and (probably) our Theorem 4.10 can be obtained by the general estimates due to Nair [32] and Nair-Tenenbaum [33] for sums $\sum_n \lambda(|P(n)|)$ with P a (primitive, possibly multi-variate) 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 [45]).

We refer to [34, Rmk 3.11] for a further discussion of variations on the $d = 1$ case of Theorem 4.10 that may be derived from other works and particularly their applicability to QUE in the level aspect.

Proof of Theorem 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 4.9 and Theorem 4.10 show that

$$(67) \qquad S_{\chi_\infty}(l, n, Y) \ll_{A,\varepsilon} X^{-A} + \sum_{r=0}^\infty 2^{-rdA} \frac{\log(2^r eU)^{d-1} 2^{rd} X}{\log(2^{rd}X)^{2-\varepsilon}} \prod_{\mathrm{N}(\mathfrak{p}) \leq 2^r X} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right).$$

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

$$(68) \qquad S_\phi(Y) \ll_{\phi,\varepsilon} \frac{k^{\mathbf{1}} Y^{3/2+\varepsilon}}{\log(k^{\mathbf{1}})^{2-\varepsilon'}} \prod_{\mathrm{N}(\mathfrak{p}) \leq k^{\mathbf{1}}} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right),$$

since the sum over l in Definition 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 (22); 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

$$(69) \qquad S_\chi(Y) \ll_\varepsilon \frac{k^{\mathbf{1}} Y^{3/2+\varepsilon}}{\log(k^{\mathbf{1}})^{2-\varepsilon'}} \prod_{\mathrm{N}(\mathfrak{p}) \leq k^{\mathbf{1}}} \left(1 + \frac{2|\lambda_f(\mathfrak{p})|}{\mathrm{N}(\mathfrak{p})}\right),$$

where we emphasize that the implied constant does not depend upon χ. By Proposition 4.2 and the definitions (31)–(32) of $M_f(x)$ and $R_f(x)$, we deduce for ϕ a Maass eigencuspform that

$$(70) \qquad \frac{\mu_f(\phi)}{\mu_f(1)} \ll_{\phi,\varepsilon} Y^{1/2+\varepsilon} \log(k^{\mathbf{1}})^{\varepsilon'} M_f(k^{\mathbf{1}})$$

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

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

The integral in (71) converges by the rapid decay of $\Psi^{\wedge}$ (see §2.6). Choosing (as Holowinsky does) $Y = \max(1, M_f(k^{\mathbf{1}})^{-1}) \ll \log(k^{\mathbf{1}})^{O(1)}$ in (70) and (71), we conclude the proof of Theorem 3.1. $\qquad\square$

5. Reduction to shifted sums weighted by an integral

In this section we establish Proposition 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 non-dihedral 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 4.1 so that

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

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 §2.8.4 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 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

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

of Mellin inversion, Cauchy's theorem and our normalization (72). We need a crude bound of the form

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

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 (74) 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 §2.8.3, shows that

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

where for simplicity we write $\kappa_{it,\infty} := \kappa_{|.|^{it},\infty}$ and $\lambda_{it} := \lambda_{|.|^{it}}$. The straightforward analysis of [46, §3.6] applied to $\zeta_{\mathbb{F}}$ in place of $\zeta_{\mathbb{Q}}$ shows that[4]

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

---

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

and it is noted in [15, 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 $|n^{\mathbf{1}}|y^{\mathbf{1}} \gg 1$)

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

Take $A = 2$. We have

$$(76) \qquad \sum_{n\in\mathbb{F}^*\cap\mathfrak{z}_j}\frac{|n^{\mathbf{1}}|^\varepsilon}{(\mathbf{max}(\mathbf{1},|n|y)^{\mathbf{1}})^2} \ll (y^{\mathbf{1}})^{-2}$$

because the LHS of (76) 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 (76) 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 bound (74).

By the rapid decay of $h^\wedge$ and the identity $h_Y^\wedge(s) = Y^s h^\wedge(s)$, we deduce from (74) that the error term in (73) 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 $Y\mu_f(1)$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

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 $i,j \in \{1,\ldots,[\mathbb{F}:\mathbb{Q}]\}$. Write the Fourier expansions of ϕ and f in the form

$$(77) \qquad \phi = \sum_{l\in\mathbb{F}}\phi_l, \quad f = \sum_{n\in\mathbb{F}^*}f_n,$$

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

$$(78) \qquad \mathcal{S}_0 = \sum_{j=1}^{h(\mathbb{F})}\int_{y\in[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]}\frac{h_Y(y^{\mathbf{1}}N(\mathfrak{z}_j))}{N(\mathfrak{z}_j)}\int_{x\in\mathbb{F}\backslash\mathbb{A}}(\phi_0|f|^2)(n(x)a(y\times z_j^{-1}))\,dx\,\frac{d^\times y}{y^{\mathbf{1}}};$$

for ϕ a Maass eigencuspform,

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

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

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

and

$$(79) \qquad |\mathcal{S}_2| \leq \mu_f(E(h_Y,\cdot))\sum_{j=1}^{h(\mathbb{F})}\sup_{\substack{y\in[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]\\h_Y(y^{\mathbf{1}}N(\mathfrak{z}_j))\neq 0}}\sum_{\substack{l\in\mathfrak{z}_j\\|l^{\mathbf{1}}|\geq Y^{1+\varepsilon}}}|\phi_l(a(y\times z_j^{-1}))|.$$

The shifted sums $S_\phi(Y)$ and $S_\chi(Y)$ are as in Definition 4.1.

---

[5]For a Maass eigencuspform, this is well known [22, Prop 10.7]; an incomplete Eisenstein series vanishes off a compact subset of $\mathbf{X}$.

Proof. By the formula (13) for integration over $Z(\mathbb{A})B(\mathbb{F})\backslash G(\mathbb{A})$, we see that

$$\mu_f(E(h_Y,\cdot)\phi)$$

$$(80) \qquad = \sum_{j=1}^{h(\mathbb{F})} \int_{y\in\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*} \frac{h_Y(y^{\mathbf{1}}\,\mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} \int_{x\in\mathbb{F}\backslash\mathbb{A}} (\phi|f|^2)(n(x)a(y\times z_j^{-1}))\,dx\,\frac{d^\times y}{y^{\mathbf{1}}}.$$

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 (80) 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 (79) follows from the formula for $\mu_f(E(h_Y,\cdot))$ given by (80) 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

$$(81) \qquad \mathcal{S}_1 = \sum_{j=1}^{h(\mathbb{F})} \sum_{\substack{(l,n)\in(\mathbb{F}^*\cap\mathfrak{z}_j)^2 \\ l^{\mathbf{1}}<Y^{1+\varepsilon} \\ n\in\mathbb{F}_{\infty+}^* \\ m:=n+l\in\mathbb{F}_{\infty+}^*}} \int_{y\in[\mathbb{F}_{\infty+}^*/\mathfrak{o}_+^*]} \frac{h_Y(y^{\mathbf{1}}\,\mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} (\phi_l\overline{f_m}f_n)(a(y\times z_j^{-1}))\,\frac{d^\times y}{y^{\mathbf{1}}}.$$

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 §2.7), 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 $\mathbb{F}_{\infty+}^*$:

$$(82) \qquad \mathcal{S}_1 = \sum_{j=1}^{h(\mathbb{F})} \sum_{\substack{(l,n)\in\mathfrak{o}_+^*\backslash(\mathbb{F}^*\cap\mathfrak{z}_j)^2 \\ l^{\mathbf{1}}<Y^{1+\varepsilon} \\ n\in\mathbb{F}_{\infty+}^* \\ m:=n+l\in\mathbb{F}_{\infty+}^*}} \int_{y\in\mathbb{F}_{\infty+}^*} \frac{h_Y(y^{\mathbf{1}}\,\mathrm{N}(\mathfrak{z}_j))}{\mathrm{N}(\mathfrak{z}_j)} (\phi_l\overline{f_m}f_n)(a(y\times z_j^{-1}))\,\frac{d^\times y}{y^{\mathbf{1}}}.$$

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 §2.8.1, §2.8.2 and §2.8.4 and the definitions of $S_\phi(Y)$ and $S_\chi(Y)$, we obtain the claimed expressions for $\mathcal{S}_1$. □

Lemma 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 (28) that

$$(83) \qquad \phi_0(y\times z_j^{-1}) = \frac{\mu(\phi)}{\mu(1)} + \sum_{1\neq\chi_0\in\mathfrak{X}(C_{\mathbb{F}}/\hat{\mathfrak{o}}^*)[2]} c_\Psi(\chi_0)\chi_0(y\times z_j^{-1}) + O_\phi(Y^{-1/2}).$$

We have

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

$$(84)$$

$$= \mu_f(E(h_Y,\cdot)) = \int_{(2)} h_Y^\wedge(s)\mu_f(E(s,\cdot))\,\frac{ds}{2\pi i},$$

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

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

Substituting (83) into (78) and applying (84) and (85), we obtain

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

Shift the contours in (86) 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 non-dihedral. Thus

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

To simplify the error term, we apply the formula

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

which follows from the unfolding method and analytic continuation as in §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 [43, p7] when $\mathbb{F} = \mathbb{Q}$), we deduce that the error term in (87) satisfies

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

with R_f given by (32). The lemma follows upon dividing through by $Y \mu_f(1)$. $\qquad \square$

Lemma 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 (79) satisfies

$$
(90) \quad 2^r (Y^{1+\varepsilon})^{1/d} \leq \max(|l_1|, \ldots, |l_d|) < 2^{r+1} (Y^{1+\varepsilon})^{1/d}
$$

for some nonnegative integer r. More generally, there are $\ll 2^{rd} Y^{1+\varepsilon}$ elements $l \in \mathfrak{z}_j$ for which (90) 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 $i, j \in \{1, \ldots, [\mathbb{F} : \mathbb{Q}]\}$, thus

$$
(91) \quad y_i \asymp Y^{-1/d} \quad \text{for each } i.
$$

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 $\kappa_{\phi,\infty}(ly) = \prod_{i=1}^d \kappa_{\phi,\infty_i}(l_i y_i)$ with

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

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

$$(92) \qquad K_{ir}(x) \ll \left(\frac{1+|r|}{x} \right)^{A'} \quad \text{uniformly for } r \in \mathbb{R} \cup i(-\tfrac{1}{2}, \tfrac{1}{2}) \text{ and } x \geq \delta > 0.$$

Thus if $l \in \mathfrak{z}_j$ and $y \in \mathbb{F}_{\infty+}^*$ satisfy (90)–(91), we obtain

$$(93) \qquad |\phi_l(a(y \times z_j^{-1}))| \ll (1 + |r|^{\mathbf{1}})^{O(1)} (2^r Y^{\varepsilon/d})^{-A}$$

for any positive A. The dependence of the bound (93) on ϕ is polynomial in the archimedean parameters r_i, so (93) extends to the case that $\phi = E(\Psi, \cdot)$ is an incomplete Eisenstein series by the integral formula (29) for its Fourier coefficients and the rapid decay of the test function $\Psi^\wedge$.

Taking A sufficiently large in (93) and summing over $l \in \mathfrak{z}_j$ that satisfy the condition (90) for some $r \in \mathbb{Z}_{\geq 0}$, we deduce

$$(94) \qquad |\mathcal{S}_2| \ll Y^{-12} \mu_f(E(h_Y, \cdot)).$$

The function h is bounded, so

$$(95) \qquad E(h_Y, g) = \sum_{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F})} h(Y|y(\gamma g)|) \ll \#\{\gamma \in B(\mathbb{F}) \backslash G(\mathbb{F}) : |y(\gamma g)| \asymp Y^{-1}\}.$$

By [47, Lem 8.7], the cardinality on the RHS of (95) 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 (94) gives $|\mathcal{S}_2| \ll Y^{-10} \mu_f(1)$. $\qquad\square$

Proof of Proposition 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 (30). $\qquad\square$

Remark 5.5. Let us point out the essential difference between our method and that of Marshall [30]. Recall that starting from Lemma 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 (29). Then

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

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

(96)
$$
\begin{aligned}
\mu_f(E(H,\cdot)\phi) = \sum_{\substack{(l,n)\in(\mathbb{F}^*\cap\mathfrak{o})^2 \\ m:=n+l\in\mathbb{F}^*\cap\mathfrak{o}}}\frac{\lambda_\phi(l)\lambda_f(m)\lambda_f(n)}{\sqrt{|l^\mathbf{1}m^\mathbf{1}n^\mathbf{1}|}} \\
\times\prod_{j=1}^{[\mathbb{F}:\mathbb{Q}]}\int_{y\in\mathbb{R}_+^*}h_0(Yy)\kappa_{\phi,\infty_j}(l_jy)\kappa_{f,\infty_j}(m_jy)\kappa_{f,\infty_j}(n_jy)\,\frac{d^\times y}{y}.
\end{aligned}
$$

The integrals here, which may be treated either by bounding κ_{ϕ,∞_j} trivially as in (54) (which is basically what Holowinsky and Marshall do) or by our sharp refinement given in Lemma 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.

6. Bounds for special functions

In this self-contained section we establish the technical lemmas that were needed in the proof of Lemma 4.3. First, recall [50] 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

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

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

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

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

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

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

Lemma 6.1. *Let $x\in\mathbb{R}_{\geq0}$, $\nu\in i\mathbb{R}\cup(-1/2,1/2)$ and $s\in\mathbb{C}$ with $\mathrm{Re}(s)\geq1/2$. Then*

$$
\left|{}_2F_1\left(\begin{matrix}\frac{1}{2}+\nu,\frac{1}{2}-\nu\\s\end{matrix};-x\right)\right|\leq1.
$$

Proof. Fix ν and s as above, and let

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

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

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

Note that since $\overline{\{\frac{1}{2} + ir, \frac{1}{2} - ir\}} = \{\frac{1}{2} + ir, \frac{1}{2} - ir\}$, 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

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

By the differential equation (97), we have

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

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 6.2. *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

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

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

$$H' = \left(F_{\nu,s}'F_{\bar{\nu},\bar{s}} + F_{\nu,s}F_{\bar{\nu},\bar{s}}'\right)\left(1 - f\frac{\lambda}{x(1 - x)}\right)$$

$$+ |F_{\nu,s}'|^2\left(f' - f\frac{2\sigma + 2u - 2(2u + 1)x}{x(1 - x)}\right).$$

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 (99). $\square$

Remark 6.3. The proof of Lemma 6.2 shows that the hypothesis $\operatorname{Re}(s) \geq 1$ can be relaxed to $\operatorname{Re}(s) \geq \frac{1}{2} + \operatorname{Re}(\nu)$; we believe that Lemma 6.2 holds in the larger range $\operatorname{Re}(s) \geq \frac{1}{2}$, $\nu \in i\mathbb{R} \cup (-\frac{1}{2}, \frac{1}{2})$, but have not proven this. Such refinements are not necessary for our applications in the proof of Lemma 4.3.

Remark 6.4. The bounds asserted by Lemmas 6.1 and 6.2 are sharp for several extremal cases of the parameters.

7. BOUNDS FOR SHIFTED SUMS UNDER HYPERBOLAS

7.1. **Reduction to a sieving problem.** In this section we establish Theorem 4.10, 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 §2.3). 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

$$(100) \qquad \Sigma_\lambda(\mathfrak{z}, l, T, U) := \sum_{\substack{n \in \mathfrak{z} \\ m := n+l \in \mathfrak{z} \\ \mathbf{max}(m,n) \in \mathcal{R}_{T,U}}} |\lambda(\mathfrak{z}^{-1}m)\lambda(\mathfrak{z}^{-1}n)|.$$

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 [4, §27], [20, p180] and [24]), 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 7.1. *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 $\operatorname{vol}(\mathcal{R}_\alpha) = X$ 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 $0 \leq x_1 \leq T_1 T_2 \cdots T_d / x_2 \cdots x_d \leq 2^{a_2 + \cdots + a_d} T_1$, 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

(101) $\qquad\qquad a_i \leq \lceil \log_2 U \rceil \text{ for } i = 2, \ldots, d \quad \text{and} \quad a_2 + \cdots a_d \geq 0.$

There are $\ll \log(eU)^{d-1}$ tuples $(a_2, \ldots, a_d) \in \mathbb{Z}^{d-1}$ satisfying the conditions (101). $\qquad\square$

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 7.1 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

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

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 $\max(b_1, \ldots, b_d) \ll X^{1/d}$, 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 4.10 reduces to the following result, which we shall establish in the remainder of this section.

Theorem 7.2. *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*

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

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

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

Preserve the hypotheses and notation of Theorem 7.2. 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 7.2 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

(105) $\qquad\qquad\qquad\qquad\qquad z = X^{1/s}, \quad s \in \mathbb{R}_{>0},$

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

$$\text{(106)} \qquad \mathfrak{z} = \sqcup \{\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} : (\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}\}.$$

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

$$\text{(107)} \qquad y = X^{\alpha}, \quad \alpha \in \mathbb{R}_{>0}$$

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 7.3; the former case will be addressed by Lemma 7.4.

Lemma 7.3. *Suppose that $2 \leq z \leq y \leq X$ with s and α as in* (105), (107) *such that $s \asymp \log\log(X)$ and $\alpha \asymp 1$. Then*

$$\text{(108)} \qquad \sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{>y}} \sum_{\substack{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n \in \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll X \log(X)^{-A}.$$

Proof. The LHS of (108) 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 (108) 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 [20, §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

$$\text{(109)} \qquad \sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{>y}} \sum_{\substack{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n \in \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll X \log(X)^{O(1)} \left(\sum_{\substack{y < |\mathfrak{a}| \leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}| \leq z}} \frac{1}{|\mathfrak{a}|} \right)^{1/4}.$$

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

$$\text{(110)} \qquad \sum_{\substack{y < |\mathfrak{a}| \leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}| \leq z}} \frac{1}{|\mathfrak{a}|} = \frac{\Psi(X, z)}{X} - \frac{\Psi(y, z)}{y} + \int_y^X \frac{\Psi(t, z)}{t^2} \, dt.$$

A theorem of Krause [25] (see also the survey [11]) 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 (110) that

$$(111) \qquad \sum_{\substack{y < |\mathfrak{a}| \leq X \\ \mathfrak{p}|\mathfrak{a} \implies |\mathfrak{p}| \leq z}} \frac{1}{|\mathfrak{a}|} \ll_A \log(X)^{-A}.$$

We deduce the required bound by substituting (111) into (109) and taking A sufficiently large. $\qquad\square$

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 7.8, set

$$(112) \qquad B(y,z) := \sup_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{\leq y}} \frac{\#\{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} : m, n \in \mathcal{R}_X\}}{\dfrac{|\mathfrak{z}^{-1}l|}{|\mathfrak{c}|^2\phi(\mathfrak{a}\mathfrak{b}\mathfrak{c}^{-1}\mathfrak{z}^{-1}l)}},$$

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

Lemma 7.4. *For y, z as in (105), (107), we have*

$$(113) \qquad \sum_{(\mathfrak{a},\mathfrak{b},\mathfrak{c}) \in \mathcal{Z}_{\leq y}} \sum_{\substack{n \in \mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \\ m,n \in \mathfrak{z} \cap \mathcal{R}_{X,\mathfrak{z}}}} \lambda^0(m)\lambda^0(n) \ll 4^s B(y,z)\log(X)^\varepsilon \prod_{|\mathfrak{p}| \leq z}\left(1 + \frac{2\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right).$$

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 $\mathfrak{z}^{-1}n = \mathfrak{b}\mathfrak{c}\mathfrak{n}$, we find similarly that $\lambda(\mathfrak{n}) \leq 2^s$. Since $\gcd(\mathfrak{a}\mathfrak{c}, \mathfrak{m}) = \gcd(\mathfrak{b}\mathfrak{c}, \mathfrak{n}) = \mathfrak{o}$, we obtain $\lambda^0(m)\lambda^0(n) = \lambda(\mathfrak{a}\mathfrak{c})\lambda(\mathfrak{b}\mathfrak{c})\lambda(\mathfrak{m})\lambda(\mathfrak{n}) \leq 4^s\lambda(\mathfrak{a}\mathfrak{c})\lambda(\mathfrak{b}\mathfrak{c})$. By the definition of $B(y,z)$ and the inequality $\phi(\mathfrak{a}\mathfrak{b}) \leq \phi(\mathfrak{a})\phi(\mathfrak{b})$, the LHS of (113) is thus

$$(114) \qquad \leq 4^s B(y,z) \sum_{\substack{\mathfrak{c}|\mathfrak{z}^{-1}l \\ \mathfrak{p}|\mathfrak{c} \implies |\mathfrak{p}| \leq z}} \frac{|\mathfrak{z}^{-1}l|}{\phi(\mathfrak{c}^{-1}\mathfrak{z}^{-1}l)|\mathfrak{c}|^2} \sum_{\substack{|\mathfrak{a}\mathfrak{c}| \leq y \\ \mathfrak{p}|\mathfrak{a}\mathfrak{b} \implies |\mathfrak{p}| \leq z}}\sum_{|\mathfrak{b}\mathfrak{c}| \leq y} \frac{\lambda(\mathfrak{a}\mathfrak{c})\lambda(\mathfrak{b}\mathfrak{c})}{\phi(\mathfrak{a})\phi(\mathfrak{b})}.$$

For $\mathfrak{c}$ as in (114), the multiplicativity of λ and ϕ implies that

$$(115) \qquad \sum_{\substack{|\mathfrak{a}\mathfrak{c}| \leq y \\ \mathfrak{p}|\mathfrak{a}\mathfrak{b} \implies |\mathfrak{p}| \leq z}}\sum_{|\mathfrak{b}\mathfrak{c}| \leq y} \frac{\lambda(\mathfrak{a}\mathfrak{c})\lambda(\mathfrak{b}\mathfrak{c})}{\phi(\mathfrak{a})\phi(\mathfrak{b})} \leq \left(\prod_{|\mathfrak{p}| \leq z}\sum_{k \geq 0} \frac{\lambda(\mathfrak{p}^{k+v_{\mathfrak{p}}(\mathfrak{c})})}{\phi(\mathfrak{p}^k)}\right)^2,$$

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

$$(116) \qquad \sum_{k \geq 0} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} = \left(1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right) \left(1 + \frac{\frac{\lambda(\mathfrak{p})}{\phi(\mathfrak{p})} - \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|} + \sum_{k \geq 2} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)}}{1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}}\right).$$

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

$$\frac{\lambda(\mathfrak{p})}{\phi(\mathfrak{p})} - \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|} + \sum_{k \geq 2} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} \leq \frac{2}{q(q-1)} + \sum_{k \geq 2} \frac{k+1}{q^{k-1}(q-1)}$$

$$= q^{-2}\left(2(1-q^{-1})^{-1} + 2(1-q^{-1})^{-2} + (1-q^{-1})^{-3}\right)$$

$$\leq 20q^{-2},$$

so that (116) implies

$$(117) \qquad \sum_{k \geq 0} \frac{\lambda(\mathfrak{p}^k)}{\phi(\mathfrak{p}^k)} \leq \left(1 + \frac{\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right)\left(1 + \frac{20}{|\mathfrak{p}|^2}\right).$$

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

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

$$= 1 + \nu\left(1 + q^{-1}(1-q^{-1})^{-2}\right) + q^{-1}(1-q^{-1})^{-2}$$

$$\leq 3\nu + 3.$$

Substituting these bounds into (114) and (115), the LHS of (113) is

$$(118) \qquad \begin{aligned} &\ll 4^s B(y,z)\psi(\mathfrak{z}^{-1}l) \prod_{|\mathfrak{p}| \leq z} \left(1 + \frac{2\lambda(\mathfrak{p})}{|\mathfrak{p}|}\right), \\ &\text{with } \psi(\mathfrak{a}) := |\mathfrak{a}| \sum_{\mathfrak{c}|\mathfrak{a}} \frac{\prod_{\mathfrak{p}^\nu||\mathfrak{c}}(3\nu+3)^2}{\phi(\mathfrak{a}/\mathfrak{c})|\mathfrak{c}|^2}. \end{aligned}$$

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

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

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 (113) is zero; if otherwise $|\mathfrak{z}^{-1}l| \leq X$, then $\psi(\mathfrak{z}^{-1}l) \ll \log(X)^\varepsilon$. Thus (113) follows from (118). $\qquad \square$

By Lemma 7.3 and Lemma 7.4, we see that Theorem 4.10 follows from sufficiently strong bounds for the quantity $B(y,z)$ given by (112); the following lemma reduces such bounds to a classical sieving problem.

Definition 7.5. For a region $\mathcal{R} \subset \mathbb{F}_\infty \cong \mathbb{R}^d$, an ideal $\mathfrak{r} \subset \mathbb{F}$, 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 $\Omega_{\mathfrak{p}} \subset \mathfrak{r}/\mathfrak{p}\mathfrak{r}$, define the *sifted set*

$$(119) \qquad \mathcal{S}(\mathcal{R}, \mathfrak{r}, (\Omega_{\mathfrak{p}})) := \{n \in \mathfrak{r} \cap \mathcal{R} : n \notin \Omega_{\mathfrak{p}} \quad (\mathfrak{p}\mathfrak{r}) \text{ for all } \mathfrak{p} \in \mathcal{P}\}.$$

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

$$(120) \qquad H((\Omega_{\mathfrak{p}}), Q) = \sum_{\substack{|\mathfrak{q}| \leq Q \\ \mathfrak{p}|\mathfrak{q} \implies \mathfrak{p} \in \mathcal{P}}} \prod_{\mathfrak{p}|\mathfrak{q}} \frac{\#\Omega_{\mathfrak{p}}}{|\mathfrak{p}| - \#\Omega_{\mathfrak{p}}}.$$

Lemma 7.6. *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*

$$\mathcal{R}_r = \{x - r \mid x \in \mathcal{R}_{X,\mathfrak{z}}\}. \tag{121}$$

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 $|\mathfrak{p}| \leq z$. 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

$$\#\Omega_\mathfrak{p} := \begin{cases} 1 & \mathfrak{p} \mid \mathfrak{a}\mathfrak{b}\mathfrak{c}^{-1}\mathfrak{z}^{-1}l \\ 2 & \text{otherwise} \end{cases} \tag{122}$$

and

$$\#(\mathfrak{z}_{\mathfrak{a},\mathfrak{b},\mathfrak{c}} \cap \mathcal{R}_{X,\mathfrak{z}}) \leq \#\mathcal{S}(\mathcal{R}_r, \mathfrak{x}, (\Omega_\mathfrak{p})). \tag{123}$$

Proof. Indeed, let $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}$, so that $\mathfrak{c} \mid \mathfrak{z}^{-1}l$ 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 $\mathfrak{p} \mid \mathfrak{a}$, $\mathfrak{p} \nmid \mathfrak{b}$. 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 $\mathfrak{p} \nmid \mathfrak{a}$, $\mathfrak{p} \mid \mathfrak{b}$, then we may similarly take $\#\Omega_\mathfrak{p} = 1$.
- The case $\mathfrak{p} \mid \mathfrak{a}, \mathfrak{p} \mid \mathfrak{b}$ 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 7.7. *Let $\mathfrak{x}$, $\mathcal{P}$, and $(\Omega_\mathfrak{p})_{\mathfrak{p} \in \mathcal{P}}$ be as in Definition 7.5. Let $\mathcal{R}$ be the region $\mathcal{R}_{X,\mathfrak{x}}$ as in (103). 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*

$$\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_\mathfrak{p})) \ll \frac{X + Q^2}{H((\Omega_\mathfrak{p}), Q)}. \tag{124}$$

Proof. See §7.2. □

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

Corollary 7.8. *Let $c_2(\mathbb{F}) > 0$ be as in Proposition 7.7. Then for $X > c_2(\mathbb{F})y^2$, the quantity $B(y, z)$ given by (112) satisfies*

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

Proof. Let $(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) \in \mathcal{Z}_{\leq y}$ and let the region $\mathcal{R}_r$, the ideal $\mathfrak{x} = \mathfrak{abc}\mathfrak{z}$, the set of primes $\mathcal{P}$ and the collection of sets of residue classes $(\Omega_{\mathfrak{p}})$ be as in Lemma 7.6, so that (123) 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 7.7 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 (122)). 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 [9, 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 (112) 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. $\qquad\square$

Proof of Theorem 7.2. Let y, z be given by (105), (107) 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 7.3, Lemma 7.4 and Corollary 7.8 are eventually satisfied, so we obtain

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

We have $4^s = \log(X)^{\alpha \log(4)}$, $\log(z) \gg_{\alpha} \log(X)^{2-\varepsilon}$ and $y^2 z^2 \ll_{\alpha} X$, so letting $\alpha \to 0$ we deduce the assertion of Theorem 7.2. $\qquad\square$

7.2. Appendix: proof of Proposition 7.7.

Inequalities of the shape (124) (with explicit constants) have appeared in papers of Schaal [38, Thm 5] and Hinz [12, 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 [4] and Kowalski [24], 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 $\alpha : \mathfrak{x}/\mathfrak{qx} \to \mathbb{C}$ a function on the group $\mathfrak{x}/\mathfrak{qx}$. Define $L^2(\mathfrak{x}/\mathfrak{qx})$, $\|.\|_2$ with respect to the counting measure, and for ψ in the Pontryagin dual $(\mathfrak{x}/\mathfrak{qx})^{\wedge}$, define $\alpha^{\wedge}(\psi) = \sum_{\mathfrak{x}/\mathfrak{qx}} \alpha(\zeta)\overline{\psi}(\zeta)$; then the Fourier inversion and Plancherel formulae read

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

For a proper divisor $\mathfrak{q}'$ of $\mathfrak{q}$, the projection $\mathfrak{x}/\mathfrak{qx} \to \mathfrak{x}/\mathfrak{q}'\mathfrak{x}$ induces an inclusion $L^2(\mathfrak{x}/\mathfrak{q}'\mathfrak{x}) \hookrightarrow L^2(\mathfrak{x}/\mathfrak{qx})$. Let $L^2_\#(\mathfrak{x}/\mathfrak{qx})$ denote the orthogonal complement of the span of the images of these inclusions, write $L^2(\mathfrak{x}/\mathfrak{qx}) \ni$

$\alpha \mapsto \alpha_\# \in L^2_\#(\mathfrak{x}/\mathfrak{q}\mathfrak{x})$ 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 $a[\mathfrak{q}](\zeta) = \sum_{n \equiv \zeta (\mathfrak{q}\mathfrak{x})} a_n$. Let $E(\cdot; \mathfrak{x}, Q)$ be the quadratic form on $V(\mathcal{R}, \mathfrak{x})$ defined by

$$(125) \qquad E((a_n); \mathfrak{x}, Q) = \sum_{\mathfrak{q} \in \mathcal{Q}} |\mathfrak{q}| \|a[\mathfrak{q}]_\#\|_2^2 = \sum_{\mathfrak{q} \in \mathcal{Q}} \sum_{(\mathfrak{x}/\mathfrak{q}\mathfrak{x})^\wedge_\#} |a[\mathfrak{q}]^\wedge(\psi)|^2,$$

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 ($\mathfrak{p}$) 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 [31] 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 [24, Lem 2.7], shows that $h(\mathfrak{q}) \|a[\mathfrak{o}]\|_2^2 \leq |\mathfrak{q}| \|a[\mathfrak{q}]_\#\|_2^2$, so recalling from (120) 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

$$(126) \qquad \#\mathcal{S}(\mathcal{R}, \mathfrak{x}, (\Omega_\mathfrak{p})) \leq \frac{D(\mathcal{R}, \mathfrak{x}, Q)}{H((\Omega_\mathfrak{p}), Q)}.$$

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

$$(127) \qquad E((a_n); \mathfrak{x}, Q) = \sum_{\xi \in \mathcal{F}(\mathfrak{x}, Q)} \left| \sum_n a_n e(\mathrm{Tr}\, \xi n) \right|^2.$$

Write $D(\mathcal{R}, \mathfrak{r}, \mathcal{F})$ synonymously for $D(\mathcal{R}, \mathfrak{r}, Q)$. The group $\mathfrak{o}_+^*$ acts on $\mathbb{F}_\infty$ and $\mathbb{F}_\infty / \mathfrak{r}^{-1} \mathfrak{d}^{-1}$ by multiplication, stabilizing $\mathfrak{r}$ 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 $\mathbb{F}_\infty / \mathfrak{r}^{-1} \mathfrak{d}^{-1}$ by the formula $d(\xi, \eta) := \min_{n \in \mathfrak{r}^{-1} \mathfrak{d}^{-1}} d_{\mathbb{F}_\infty}(\xi, \eta + n)$ 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 7.9. $\delta(\mathcal{F}(\mathfrak{r}, Q)) \geq (|\mathfrak{r}| \Delta_\mathbb{F} Q^2)^{-1/[\mathbb{F}:\mathbb{Q}]}$ *(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. $\square$

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

$$(128) \qquad \sum_{n \in \mathfrak{r} \cap \mathcal{R}} \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2 \leq D(\mathcal{R}, \mathfrak{r}, \mathcal{F}) \|b\|_2^2$$

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 (128) and invoking the elementary inequality $|b_\xi \overline{b_\eta}| \leq \frac{1}{2}(|b_\xi|^2 + |b_\eta|^2)$, we find that

$$\sum_{n \in \mathfrak{r} \cap \mathcal{R}} \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2 \leq \sum_{n \in \mathfrak{r}} f(n) \left| \sum_{\xi \in \mathcal{F}} b_\xi e(\operatorname{Tr} \xi n) \right|^2$$

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

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

$$\sum_{n \in \mathfrak{r}} f(n) e(\operatorname{Tr} n(\xi - \eta)) = \operatorname{vol}(\mathbb{F}_\infty / \mathfrak{r})^{-1} \sum_{\mu \in \mathfrak{r}^{-1} \mathfrak{d}^{-1}} \hat{f}(\mu - \xi + \eta),$$

$$\text{with } \hat{f}(y) := \int_{\mathbb{F}_\infty} f(x) e(-x \cdot y) \, dy,$$

we obtain

$$D(\mathcal{R}, \mathfrak{r}, \mathcal{F}) \leq \operatorname{vol}(\mathbb{F}_\infty / \mathfrak{o})^{-1} |\mathfrak{r}|^{-1} F(f; \mathfrak{r}, \mathcal{F}),$$

$$(129) \qquad \text{with } F(f; \mathfrak{r}, \mathcal{F}) := \sup_{\xi \in \mathcal{F}} \sum_{\eta \in \mathcal{F}} \left| \sum_{\mu \in \mathfrak{r}^{-1} \mathfrak{d}^{-1}} \hat{f}(\mu - \xi + \eta) \right|.$$

Lemma 7.10. *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 $\operatorname{vol}(\mathcal{R}) = \prod |a_i - b_i|$ satisfies $\operatorname{vol}(\mathcal{R}) > c_2(\mathbb{F})|\mathfrak{r}|$, there exists an $\mathcal{R}$-admissible function f such that*

$$(130) \qquad F(f, \mathfrak{r}, \mathcal{F}) \ll_\mathbb{F} \operatorname{vol}(\mathcal{R}) + \delta^{-d}.$$

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 $i, j \in \{1, \ldots, d\}$, where the implied constant depends only upon $\mathbb{F}$. Now the formula

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

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 $\text{vol}(\mathcal{R}) > c_2(\mathbb{F})|\mathfrak{r}|$ 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 (129) 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

$$(131) \qquad F(f, \mathfrak{r}, \mathcal{F}) \leq \left(\frac{\pi^2}{4}\right)^d \prod_{i=1}^d |a_i - b_i|(1 + \lfloor\delta^{-1}|c_i - d_i|\rfloor) \ll \prod_{i=1}^d (|a_i - b_i| + \delta^{-1}).$$

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

Proof of Proposition 7.7. Take $c_2(\mathbb{F})$ as in Lemma 7.10, and suppose that $X > c_2(\mathbb{F})$ and $Q \geq 1$. Then $\text{vol}(\mathcal{R}_{X,\mathfrak{z}}) > c_2(\mathbb{F})|\mathfrak{z}|$, so the hypotheses of Lemma 7.10 are satisfied. The claimed bound (124) follows immediately from (126), Lemma 7.9, equation (129) and Lemma 7.10. $\square$

References

[1] Don Blasius. Hilbert modular forms and the Ramanujan conjecture. In *Noncommutative geometry and number theory*, Aspects Math., E37, pages 35–56. Vieweg, Wiesbaden, 2006.
[2] V. Blomer and G. Harcos. Twisted L-functions over number fields and Hilbert's eleventh problem. *ArXiv e-prints*, April 2009.
[3] Y. Colin de Verdière. Ergodicité et fonctions propres du laplacien. In *Bony-Sjöstrand-Meyer seminar, 1984–1985*, pages Exp. No. 13, 8. École Polytech., Palaiseau, 1985.
[4] Harold Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1980. Revised by Hugh L. Montgomery.
[5] John Friedlander and Henryk Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.
[6] Stephen Gelbart and Hervé Jacquet. A relation between automorphic representations of GL(2) and GL(3). *Ann. Sci. École Norm. Sup. (4)*, 11(4):471–542, 1978.
[7] Stephen Gelbart and Hervé Jacquet. Forms of GL(2) from the analytic point of view. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1*, Proc. Sympos. Pure Math., XXXIII, pages 213–251. Amer. Math. Soc., Providence, R.I., 1979.
[8] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 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).

[9] George Greaves. *Sieves in number theory*, volume 43 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 2001.

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

[11] Adolf Hildebrand and Gérald Tenenbaum. Integers without large prime factors. *J. Théor. Nombres Bordeaux*, 5(2):411–484, 1993.

[12] Jürgen G. Hinz. Methoden des grossen Siebes in algebraischen Zahlkörpern. *Manuscripta Math.*, 57(2):181–194, 1987.

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

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

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

[16] Roman Holowinsky. A sieve method for shifted convolution sums. *Duke Math. J.*, 146(3):401–448, 2009.

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

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

[19] Henryk Iwaniec. Notes on the quantum unique ergodicity for holomorphic cusp forms, 2010.

[20] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[21] Henryk Iwaniec and Peter Sarnak. Perspectives on the analytic theory of *L*-functions. *Geom. Funct. Anal.*, (Special Volume, Part II):705–741, 2000. GAFA 2000 (Tel Aviv, 1999).

[22] H. Jacquet and R. P. Langlands. *Automorphic forms on* GL(2). Lecture Notes in Mathematics, Vol. 114. Springer-Verlag, Berlin, 1970.

[23] Hervé Jacquet. *Automorphic forms on* GL(2). *Part II*. Lecture Notes in Mathematics, Vol. 278. Springer-Verlag, Berlin, 1972.

[24] E. Kowalski. *The large sieve and its applications*, volume 175 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2008. Arithmetic geometry, random walks and discrete groups.

[25] Uwe Krause. Abschätzungen für die Funktion $\Psi_K(x, y)$ in algebraischen Zahlkörpern. *Manuscripta Math.*, 69(3):319–331, 1990.

[26] J.-P. Labesse and R. P. Langlands. *L*-indistinguishability for SL(2). *Canad. J. Math.*, 31(4):726–785, 1979.

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

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

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

[30] S. Marshall. Mass Equidistribution for Automorphic Forms of Cohomological Type on GL_2. *ArXiv e-prints*, June 2010.

[31] H. L. Montgomery. A note on the large sieve. *J. London Math. Soc.*, 43:93–98, 1968.

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

[33] Mohan Nair and Gérald Tenenbaum. Short sums of certain arithmetic functions. *Acta Math.*, 180(1):119–144, 1998.

[34] Paul Nelson. Equidistribution of cusp forms in the level aspect. `http://www.its.caltech.edu/~pnelson/hque-level.pdf`, 2010.

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

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

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

[38] Werner Schaal. On the large sieve method in algebraic number fields. *J. Number Theory*, 2:249–270, 1970.

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

[40] Goro Shimura. The special values of the zeta functions associated with Hilbert modular forms. *Duke Math. J.*, 45(3):637–679, 1978.

[41] Lior Silberman and Akshay Venkatesh. On quantum unique ergodicity for locally symmetric spaces. *Geom. Funct. Anal.*, 17(3):960–998, 2007.

[42] A. I. Šnirel′man. Ergodic properties of eigenfunctions. *Uspehi Mat. Nauk*, 29(6(180)):181–182, 1974.

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

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

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

[46] E. C. Titchmarsh. *The theory of the Riemann zeta-function*. The Clarendon Press Oxford University Press, New York, second edition, 1986. Edited and with a preface by D. R. Heath-Brown.

[47] Akshay Venkatesh. Sparse equidistribution problems, period bounds, and subconvexity. `arXiv.org:math/0506224`, 2005.

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

[49] André Weil. Séries de Dirichlet et fonctions automorphes. In *Séminaire Bourbaki, Vol. 10*, pages Exp. No. 346, 547–552. Soc. Math. France, Paris, 1995.

[50] 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.

[51] Steven Zelditch. Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.*, 55(4):919–941, 1987.