BOUNDS FOR MODULAR L-FUNCTIONS IN THE LEVEL ASPECT

V. BLOMER, G. HARCOS, AND P. MICHEL
MARCH 9, 2012

ABSTRACT. Let f be a primitive (holomorphic or Maaß) cusp form of level q and nontrivial nebentypus. Then for $\mathfrak{Re}\, s = \frac{1}{2}$ the associated L-function satisfies $L(f,s) \ll q^{\frac{1}{4}-\frac{1}{1889}}$, where the implied constant depends polynomially on s and the archimedean parameters of f (weight or Laplacian eigenvalue).

RÉSUMÉ. Soit f une forme modulaire cuspidale primitive (holomorphe ou de Maaß) de niveau q et de caractère central non-trivial. Alors, pour $\mathfrak{Re}\, s = \frac{1}{2}$, la function L associée vérifie $L(f,s) \ll q^{\frac{1}{4}-\frac{1}{1889}}$, où la constante implicite dépend polynômialement de s et du paramètre archimédien de f (le poids ou la valeur propre du Laplacien).

1. INTRODUCTION

1.1. **Statement of results.** It is one of the magic features of analytic continuation that L-functions reveal the most relevant information of the coefficients they encode, in a region where the Dirichlet series does not converge. For many obvious (e.g. bounding contour integrals) and not so obvious (e.g. equidistribution problems) applications it turns out to be crucial to have estimates for the size of L-functions inside the critical strip, and without much loss of generality on the critical line $\mathfrak{Re}\, s = \frac{1}{2}$. Typical L-functions come equipped with a functional equation which, by the standard Phragmén–Lindelöf convexity principle, implies an upper bound in the critical strip, the so-called convexity bound. It turns out, however, that for many of the deeper questions that are attacked by an analytic machinery, one needs to improve on this convexity bound. Interestingly, even (seemingly) marginal improvements often result in quite strong applications (the reader may consult the surveys [Fr, IS, MV]). In an impressive series of papers, Duke–Friedlander–Iwaniec developed powerful methods to obtain bounds that could break the convexity barrier for various families of L-functions, culminating in a subconvexity bound for general automorphic forms on GL_2 [DFI5]:

Theorem 1 (Duke–Friedlander–Iwaniec)**.** *Let f be a primitive holomorphic or Maaß cusp form of level q, with primitive nebentypus χ, and archimedean parameter t_f. Then for $\mathfrak{Re}\, s = \frac{1}{2}$ the associated L-function satisfies*

$$L(f,s) \ll (|s| + |t_f|)^{\frac{19}{2}} q^{\frac{1}{4}-\frac{1}{23041}}.$$

The convexity bound in this context is $L(f,s) \ll_{s,t_f,\varepsilon} q^{\frac{1}{4}+\varepsilon}$. Here and below the archimedean parameter t_f of f is defined as

$$t_f := \begin{cases} \sqrt{\lambda_f - \frac{1}{4}} & \text{when } f \text{ is a Maaß form of Laplacian eigenvalue } \lambda_f, \\ \frac{(1-k)i}{2} & \text{when } f \text{ is a holomorphic form of weight } k. \end{cases} \tag{1.1}$$

2000 *Mathematics Subject Classification.* Primary: 11F66; Secondary: 11F12.

Key words and phrases. Automorphic forms, L-series, subconvexity, amplification, additive divisor sums.

First and second authors supported by NSERC grant 311664-05 and NSF grant DMS-0503804. Third author supported by Marie Curie RT Network "Arithmetic Algebraic Geometry" and by the "RAP" network of the Région Languedoc-Roussillon.

Theorem 1 is a major breakthrough, and the proof is long and very elaborate. In this paper we want to present a different method that avoids a number of technical difficulties and gives a better exponent in a more general setting. The main result of the present paper is

Theorem 2. *Let f be a primitive[1] (holomorphic or Maaß) cusp form of level q. Suppose that the nebentypus χ of f is not trivial. Then for $\mathfrak{Re}\, s = \frac{1}{2}$ the associated L-function satisfies*

$$L(f,s) \ll (|s|+|t_f|)^A q^{\frac{1}{4}-\frac{1}{1889}},$$

where $A > 0$ is an absolute constant.

Remark 1.1. Note that besides the improved subconvex exponent, we do not require, as in [DFI5], that the nebentypus χ is primitive[2]: this is a feature that is easily allowed by our method. There is no doubt that the method of [DFI5] could—in principle—also be adapted to cover the case of non-primitive nebentypus, but—we believe—at the expense of extremely cumbersome and technical computations. Including non-primitive characters is not only a cosmetical device. Corollaries 1 and 2 below give nice applications that rest crucially on this more general setting.

Remark 1.2. With some major effort we could probably obtain $A = 2$, but here we do not focus on the exact dependence on the other parameters except that we keep it polynomial.

We shall describe our method and the new ideas briefly in the next section, but we state already at this place the most fundamental difference to the approach in [DFI5]: one of the great technical difficulties in [DFI5] was to match the contribution of the Eisenstein spectrum, added to the sum in order to make it spectrally complete, by a suitable term on the other side of the trace formula. We will be able to avoid this matching problem completely so that throughout the paper upper bounds suffice. This gives more flexibility and makes it feasible to include non-primitive characters for which explicit computations with Eisenstein series seem to be hard.

As in [Mi, HM] a crucial ingredient for Theorem 2 is a subconvex estimate on the critical line for a smaller family of L-functions, namely automorphic L-functions twisted by a character of large conductor. More precisely, we use the fact that for any primitive (holomorphic or Maaß) cusp form f of level N and trivial nebentypus and for any primitive character χ of modulus q, the twisted L-function $L(f\otimes\chi,s)$ satisfies on the critical line

$$L(f\otimes\chi,s) \ll_\varepsilon (|s|(1+|t_f|)Nq)^\varepsilon |s|^\alpha (1+|t_f|)^\beta N^\gamma q^{\frac{1}{2}-\delta} \tag{1.2}$$

for some absolute positive constants α, β, γ, δ and any $\varepsilon > 0$. In [BHM] the authors obtained (1.2) with

$$\alpha = \frac{503}{256}, \qquad \beta = \frac{1221}{256}, \qquad \gamma = \frac{9}{16}, \qquad \delta = \frac{25}{256},$$

in the somewhat more general setting where f was allowed to have any nebentypus. While any bound of type (1.2) suffices for our purposes, in order to obtain a good exponent, we cite the following result from [BH]:

Theorem 3. *Let f be a primitive Maaß cusp form of weight zero, archimedean parameter t_f, level N and trivial nebentypus, and let χ be a primitive character modulo q. Then for $\mathfrak{Re}\, s = \frac{1}{2}$ the twisted L-function satisfies*

$$L(f\otimes\chi,s) \ll_\varepsilon \left(|s|^{\frac{1}{4}}(1+|t_f|)^3 N^{\frac{1}{4}} q^{\frac{3}{8}} + |s|^{\frac{1}{2}}(1+|t_f|)^{\frac{7}{2}} N^{\frac{3}{4}} q^{\frac{1}{4}}\right)(|s|(1+|t_f|)Nq)^\varepsilon.$$

In particular,

$$L(f\otimes\chi,s) \ll_\varepsilon (|s|(1+|t_f|)Nq)^\varepsilon\, |s|^{\frac{1}{2}}(1+|t_f|)^3 N^{\frac{1}{4}} q^{\frac{3}{8}}, \tag{1.3}$$

provided $q \geqslant (N(1+|t_f|))^4$.

[1] i.e., eigenform of all Hecke operators

[2] In fact, with slightly more work we could also have covered the trivial nebentypus case (see Remark 4.1).

The proof of this theorem uses and generalizes a method of Bykovskiĭ [By]. Since exponents get improved and the numerical values of α, β, γ, δ in (1.2) might be lowered in some future, it seemed convenient to calculate the subconvex exponent in Theorem 2 as a function of these constants, see (7.5).

As in [DFI5] we derive the following corollary from Theorem 2:

Corollary 1. *Let K be quadratic field and $\mathcal{O} \subset K$ an order in K of discriminant d. Let χ denote a (primitive) character of* $\mathrm{Pic}(\mathcal{O})$. *Then for $\mathfrak{Re}\, s = \frac{1}{2}$ the associated L-function satisfies*

$$L(\chi, s) \ll_s d^{\frac{1}{4} - \frac{1}{1889}}.$$

Indeed, by theorems of Hecke and Maaß, $L(\chi, s)$ is the L-function of a Maaß form of weight $\kappa \in \{0, 1\}$ (depending on whether K is real or imaginary), level d and nebentypus χ_K (the quadratic character associated with K). One difference with Theorem 2.7 of [DFI5] is that we do not require the character χ to be associated with the maximal order $\mathcal{O}_K$. Of course a similar bound holds for any Hecke character of a quadratic field. An immediate consequence of Corollary 1 is the following:

Corollary 2. *Let K be a cubic field of discriminant d_K. Then for $\mathfrak{Re}\, s = \frac{1}{2}$ its associated Dedekind L-function satisfies*

$$\zeta_K(s) \ll_s |d_K|^{\frac{1}{4} - \frac{1}{1889}}. \tag{1.4}$$

Proof. If K is abelian, then $d_K = d^2$ is a square and $\zeta_K(s) = \zeta(s)L(\chi, s)L(\overline{\chi}, s)$, where χ is a Dirichlet character of order 3 and conductor d. In that case the bound (1.4) follows from Burgess's subconvex bound [Bu]. If K is not abelian, let L denote the Galois closure of K (which is of degree 6 with Galois group isomorphic to S_3) and let $F/\mathbb{Q}$ denote the unique quadratic field contained in L, then $\zeta_K(s) = \zeta(s)L(\chi, s)$, where χ is a ring-class character of F of order 3 and conductor $\mathfrak{d}$ satisfying $N_{F/\mathbb{Q}}(\mathfrak{d}) = d_K$. The bound (1.4) now follows from Corollary 1. □

This latter bound turns out to be an important ingredient in the work of Einsiedler, Lindenstrauss, Venkatesh and the third author [ELMV] where a higher rank generalization of Duke's equidistribution theorem for closed geodesics on the modular surface ([Du, Theorem 1]) is established.

1.2. **Outline of the Proof.** Let us give an overview of the new ideas involved and of the various steps that we perform in order to prove Theorem 2. Basically we work out the strategy outlined in [Mi, Section 1.3.1] but with various innovations which were not anticipated at that time. As the reader will notice, the methods of [KMV, DFI5, HM] are the ancestors of our proof (rather than the method followed in [Mi]). The ultimate goal is to estimate an amplified fourth moment

$$\frac{1}{q} \sum_f |M(f)|^2 |L(f, s)|^4, \qquad \mathfrak{Re}\, s = \frac{1}{2} \tag{1.5}$$

over the spectrum of the Laplacian acting on automorphic functions f of level q, nebentypus χ, and some weight $\kappa \in \{0, 1\}$. Here M is a suitable amplifier, and f runs through Maaß cusp forms and Eisenstein series (in the latter case, of course, the sum becomes an integral). As usual, the amplifier $M(f) = \sum_\ell x(\ell)\lambda_f(\ell)$ is a short Dirichlet polynomial; opening the square and using multiplicativity of Hecke eigenvalues, we are left with bounding a normalized average

$$\mathcal{Q}(\ell) := \frac{1}{q} \sum_f \lambda_f(\ell) |L(f, s)|^4$$

for ℓ less than a small power of q (see Section 3). We win once we can show $\mathcal{Q}(\ell) \ll \ell^{-\delta}$ for some $\delta > 0$ (cf. (3.20)). Using an approximate functional equation for $L(f, s)^2$, $|L(f, s)|^4$ can be written as a double sum, each of length about q, involving divisor functions and Hecke eigenvalues. Thus $|L(f, s)|^4$ can be viewed as the square of a Rankin–Selberg L-function of f times an Eisenstein series. We can therefore use ideas from [Mi, HM], but we are faced with various difficulties coming from the fact that Eisenstein series are not square-integrable; we shall discuss this below.

By Kuznetsov's trace formula, the spectral sum is transformed into sums of Kloosterman sums, twisted by χ. For one of the remaining divisor functions, we apply Voronoi summation so that the twisted Kloosterman sums become Gauß sums (cf. (4.6)). Opening the Gauß sum (as in (5.5)), we are now left with sums roughly of the type

$$\frac{1}{q^{3/2}}\sum_g \frac{1}{g}\sum_h \chi(h)\sum_{\ell n\pm m=gh}\tau(n)\tau(m)g_\ell(n,m). \tag{1.6}$$

Here the length of n, m, h is about q and the function g_ℓ is essentially bounded (precisely, see (4.7) and (4.9)). Hence we need not only square-root cancellation in the character sum, but also some additional saving in the ℓ-variable. There are several ways to evaluate the inner sum. The δ-symbol method as in [DFI1] would be one alternative; however, we shall use a related method that originally goes back to Heath-Brown and was used by Meurman [Me] in a very nice but somewhat neglected paper. The idea is to start with a smooth variant of $\tau(n) \approx 2\#\{d \mid n : d \leqslant \sqrt{n}\}$ and to transform the other divisor function $\tau(m) = \tau(\pm(h-\ell n))$ by Voronoi summation. This method is quicker and simpler than the δ-symbol method, and there are fewer error terms to take care of. This simplifies the already cumbersome estimations in Section 6 considerably. The error terms of the additive divisor problem (1.6) involve Kloosterman sums. Meurman's method has the additional advantage that the multiplicative inverse $\bar\ell$ does not enter the variables of the Kloosterman sum; its removal by switching to another cusp would be cumbersome for ℓ not square-free. We transform the Kloosterman sums into Fourier coefficients of automorphic forms f (say) of level at most ℓ by using Kuznetsov's formula in the *other* direction. Now we have smooth sums of the type

$$\sum_h \sum_f \chi(h)\rho_f(h), \tag{1.7}$$

where the h-sum has about length q. Cancellation in the h-sum is therefore equivalent to subconvexity of twisted automorphic L-functions for which we need Theorem 3. We will carry this out in Section 6.3. Some difficulties arise from the fact that (1.6) may be "ill-posed": if the support of g_ℓ is such that m is much smaller than n, we have to solve an unbalanced shifted convolution problem which is reflected by the fact that the f-sum in (1.7) is long, cf. (6.16). In this case the saving comes from the spectral large sieve inequalities of Deshouillers–Iwaniec, see Section 6.2.

The preceding discussion shows that we have presented here a fairly general method for calculating accurately twisted sums of additive divisor sums. This may have applications in different contexts.

Comparing with the approach in [DFI5], two things are different: we interpret $|L(f,s)|^4$ as $|L(f,s)^2|^2$ rather than $(|L(f,s)|^2)^2$, cf. [Mi, Section 1.3.1]. This subtle difference is mainly responsible for a better exponent in the final result. It turns out that the shifted convolution problem we are facing in the course of the proof has a nice arithmetic interpretation, and we can dispense with a general (and therefore weaker) result on fairly arbitrary determinant-type equations [DFI3]. In fact, we have removed the appearance of the character χ inside the shifted convolution problem, but the price we have to pay is that the shifted convolution sums are now twisted by χ; our approach is only successful because we can exploit some cancellation coming from the character sum.

Secondly, we tailor the approximate functional equation according to our needs. The main term in (1.6) can be calculated directly (see Section 5), and it turns out that for each fixed g, the sum is $\ll \ell^{-1/2}$, but the length of the g-sum is about $\ell^{1/2}$. Precisely, using Mellin transforms, the g-sum translates into a zeta-function (cf. (5.7)) whose residue gives a main term. The remaining integral is indeed $\ll \ell^{-1/2}$, but the polar contribution is in general $\gg 1$. We are facing exactly the same problem as in [DFI5]: the spectral sum (1.5) is too large, and by adding artificially the Eisenstein spectrum in order to make the sum spectrally complete, we have added a term whose contribution is larger than what we want to estimate. If we work in the setting of holomorphic cusp forms and use Petersson's trace formula instead of Kuznetsov's, a certain orthogonality relation between the Bessel functions J_n, $n \equiv \kappa \pmod 2$, and the functions Y_0, K_0 coming from the additive divisor problem, creates a zero that like a *deus ex machina* kills the pole from the zeta-function. In various variants,

this has been the key to success in many related papers, cf. [DFI2, (55)], [DFI4, Lemma 7.1], [KMV, Sections 3 and 4]. Since we have quite some flexibility with the weight functions of the approximate functional equation on the one hand and of the trace formula on the other hand, we were able to create such a zero artificially. The choice of the approximate functional equation below ((3.9)–(3.11)) is motivated by forcing the Eisenstein contribution to be small, see Remark 3.1. This device might also be useful in different applications.

There are other ways to avoid a large contribution of the Eisenstein spectrum. For example, when f is odd, i.e., has sign $\epsilon_f = -1$, we can insert a factor $1 - \epsilon_f$ into the trace formula which amounts to subtracting the trace formula (2.12) for $mn < 0$ from the trace formula (2.11) for $mn > 0$. Since Eisenstein series are even, their contribution vanishes—as everything else that is even—which is mirrored in the fact that the polar term in Section 5 will vanish automatically. Another approach might be to work with the finite places and encode the arithmetic information of the (Hecke) coefficients of the Eisenstein series into the amplifier.

We believe that a line of attack that totally dispenses with approximate functional equations (as in [By]) would give a cleaner and simpler proof, but we have not yet succeeded in completing this project. The main reason is that it seems hard to translate shifted convolution problems into a language without finite sums.

Acknowledgements. This paper was worked out at several places: while the three authors were invited at the workshop "Theory of the Zeta and Allied Functions" by the Mathematisches Forschungsinstitut Oberwolfach in September 2004; while the first author was invited by the Université Montpellier II in March 2005; while the second author was invited by the University of Toronto in September 2005; while the first and second authors were invited at the workshop "Gaps between Primes" by the American Institute of Mathematics in November 2005; while the first author was invited by the University of Texas at Austin in February 2006 and while the third author was invited at the Institut des Hautes Études Scientifiques from March to June 2006. We would like to thank all six institutions as well as Massey College, Toronto, for their hospitality and inspiring working conditions.

2. Preliminaries

2.1. **Automorphic forms.** In this section we briefly compile some results from the theory of automorphic forms which we shall need later. An exhaustive account of the theory can be found in [DFI5] from which we borrow much of the notation. One of the most difficult issues in this subject is the normalization. We normalize the Fourier coefficients as in [DFI5] and write the trace formulae as in [Iw, Chapter 9], i.e., using the normalization [Iw, (8.5)–(8.6)].

2.1.1. *Hecke eigenbases.* Let $q \geqslant 1$ be an integer, χ be a character to modulus q; let $\kappa = \frac{1-\chi(-1)}{2} \in \{0, 1\}$ and $k \geqslant 2$ be an integer satisfying $(-1)^k = \chi(-1)$. We denote by $\mathcal{S}_k(q, \chi)$, $\mathcal{L}^2(q, \chi)$ and $\mathcal{L}^2_0(q, \chi) \subset \mathcal{L}^2(q, \chi)$, respectively, the Hilbert spaces (with respect to the Petersson inner product) of holomorphic cusp forms of weight k, of Maaß forms of weight κ, and of Maaß cusp forms of weight κ, with respect to the congruence subgroup $\Gamma_0(q)$ and with nebentypus χ. These spaces are endowed with the action of the (commutative) algebra $\mathbf{T}$ generated by the Hecke operators $\{T_n \mid n \geqslant 1\}$. Moreover, the subalgebra $\mathbf{T}^{(q)}$ generated by $\{T_n \mid (n, q) = 1\}$ is made of normal operators. As an immediate consequence, the spaces $\mathcal{S}_k(q, \chi)$ and $\mathcal{L}^2_0(q, \chi)$ have an orthonormal basis made of eigenforms of $\mathbf{T}^{(q)}$ and such a basis can be chosen to contain all L^2-normalized Hecke eigen-newforms (in the sense of Atkin–Lehner theory). We denote these bases by $\mathcal{B}_k(q, \chi)$ and $\mathcal{B}(q, \chi)$, respectively. For the rest of this paper we assume that any such basis satisfies these properties.

The orthogonal complement to $\mathcal{L}^2_0(q, \chi)$ in $\mathcal{L}^2(q, \chi)$ is the Eisenstein spectrum $\mathcal{E}(q, \chi)$ (plus possibly the space of constant functions if χ is trivial). The space $\mathcal{E}(q, \chi)$ is continuously spanned by a "basis" of Eisenstein series indexed by some finite set. In the classical setting, that set is usually taken to be the set $\{\mathfrak{a}\}$ of cusps of $\Gamma_0(q)$ which are singular with respect to χ. In that case the

spectral decomposition of any $\psi \in \mathcal{E}(q,\chi)$ reads

$$\psi(z) = \sum_{\mathfrak{a}} \int_{\mathbb{R}} \langle \psi, E_{\mathfrak{a}}(\cdot, \tfrac{1}{2}+it)\rangle E_{\mathfrak{a}}(z, \tfrac{1}{2}+it)\frac{dt}{4\pi}.$$

Such a basis has the advantage of being explicit and indeed it will turn out to be useful at the end of our argument. On the other hand, it will be equally useful for us to employ another basis of Eisenstein series formed of Hecke eigenforms: the adelic reformulation of the theory of modular forms provides a natural spectral expansion of the Eisenstein spectrum in which the basis of Eisenstein series is indexed by a set of parameters of the form[3]

$$\{(\chi_1,\chi_2,f) \mid \chi_1\chi_2 = \chi,\ f \in \mathcal{B}(\chi_1,\chi_2)\}, \tag{2.1}$$

where (χ_1,χ_2) ranges over the pairs of characters of modulus q such that $\chi_1\chi_2 = \chi$ and $\mathcal{B}(\chi_1,\chi_2)$ is some finite set depending on (χ_1,χ_2) (specifically, $\mathcal{B}(\chi_1,\chi_2)$ corresponds to an orthonormal basis in the space of an induced representation constructed out of the pair (χ_1,χ_2), but we need not be more precise). We refer to [GJ] for the definition of these parameters as well as for the proof of the spectral expansion of this form. With this choice, the spectral expansion for $\psi \in \mathcal{E}(q,\chi)$ reads

$$\psi(z) = \sum_{\substack{\chi_1\chi_2=\chi \\ f\in\mathcal{B}(\chi_1,\chi_2)}} \sum \int_{\mathbb{R}} \langle \psi, E_{\chi_1,\chi_2,f}(\cdot, \tfrac{1}{2}+it)\rangle E_{\chi_1,\chi_2,f}(z, \tfrac{1}{2}+it)\frac{dt}{4\pi}.$$

The main advantage is that these Eisenstein series are Hecke eigenforms for $\mathbf{T}^{(q)}$: for $(n,q)=1$, one has

$$T_n E_{\chi_1,\chi_2,f}(z, \tfrac{1}{2}+it) = \lambda_{\chi_1,\chi_2}(n,t) E_{\chi_1,\chi_2,f}(z, \tfrac{1}{2}+it)$$

with

$$\lambda_{\chi_1,\chi_2}(n,t) = \sum_{ab=n} \chi_1(a)a^{it}\chi_2(b)b^{-it}.$$

2.1.2. *Multiplicative and boundedness properties of Hecke eigenvalues.* Let f be any such Hecke eigenform and let $\lambda_f(n)$ denote the corresponding eigenvalue for T_n; then for $(mn,q)=1$ one has

$$\lambda_f(m)\lambda_f(n) = \sum_{d\mid(m,n)} \chi(d)\lambda_f(mn/d^2), \tag{2.2}$$

$$\overline{\lambda_f(n)} = \overline{\chi(n)}\lambda_f(n).$$

In particular, for $(mn,q)=1$ it follows that

$$\lambda_f(m)\overline{\lambda_f(n)} = \overline{\chi(n)} \sum_{d\mid(m,n)} \chi(d)\lambda_f(mn/d^2). \tag{2.3}$$

Note also that the formula (2.2) is valid for all m, n if f is an eigenform for all $\mathbf{T}$. In particular, this is the case when $\lambda_f(n)$ is replaced by the divisor function and χ is the trivial character.

We recall the bounds satisfied by the Hecke eigenvalues: if f belongs to $\mathcal{B}_k(q,\chi)$ (i.e., is holomorphic) or is an Eisenstein series $E_{\chi_1,\chi_2,f}(z,\frac{1}{2}+it)$, then one has

$$|\lambda_f(n)| \leqslant \tau(n) \ll_\varepsilon n^\varepsilon$$

for any $\varepsilon > 0$. For $f \in \mathcal{B}(q,\chi)$ the currently best approximation due to Kim–Sarnak [KS] is

$$|\lambda_f(n)| \leqslant \tau(n) n^\theta \quad \text{with} \quad \theta := \frac{7}{64}. \tag{2.4}$$

There is an analogous bound for the spectral parameter (1.1):

$$|\Im\mathfrak{m}\, t_f| \leqslant \theta. \tag{2.5}$$

[3]We suppress here the independent spectral parameters $\frac{1}{2}+it$ with $t\in\mathbb{R}$.

Moreover, $t_f \in \mathbb{R}$ when $\kappa = 1$. [DFI5, Proposition 19.6] shows that the above bound is valid with $\theta = 0$ on average over n:

$$\sum_{n \leqslant x} |\lambda_f(n)|^2 \ll_\varepsilon ((1+|t_f|)qx)^\varepsilon x \tag{2.6}$$

for any $x \geqslant 1$, $\varepsilon > 0$.

2.1.3. *Hecke eigenvalues and Fourier coefficients.* We write the Fourier expansion of a modular form f as follows ($z = x + iy$):

$$f(z) = \sum_{n \geqslant 1} \rho_f(n) n^{k/2} e(nz) \quad \text{for} \quad f \in \mathcal{B}_k(q,\chi),$$

$$f(z) = \sum_{n \neq 0} \rho_f(n) W_{\frac{n}{|n|}\frac{\kappa}{2}, it_f}(4\pi|n|y) e(nx) \quad \text{for} \quad f \in \mathcal{B}(q,\chi),$$

(here t_f denotes the spectral parameter (1.1)) and for either type of Eisenstein series

$$E_{\mathfrak{a}}(z, \tfrac{1}{2}+it) = c_{1,\mathfrak{a}}(t) y^{1/2+it} + c_{2,\mathfrak{a}}(t) y^{1/2-it} + \sum_{n \neq 0} \rho_{\mathfrak{a}}(n,t) W_{\frac{n}{|n|}\frac{\kappa}{2}, it}(4\pi|n|y) e(nx),$$

$$E_{\chi_1,\chi_2,f}(z, \tfrac{1}{2}+it) = c_{1,f}(t) y^{1/2+it} + c_{2,f}(t) y^{1/2-it} + \sum_{n \neq 0} \rho_f(n,t) W_{\frac{n}{|n|}\frac{\kappa}{2}, it}(4\pi|n|y) e(nx).$$

When f is a Hecke eigenform, there is a close relationship between the Fourier coefficients of f and its Hecke eigenvalues $\lambda_f(n)$: one has for $(m,q) = 1$ and any $n \geqslant 1$,

$$\lambda_f(m)\sqrt{n}\rho_f(n) = \sum_{d|(m,n)} \chi(d) \sqrt{\frac{mn}{d^2}} \rho_f\left(\frac{mn}{d^2}\right); \tag{2.7}$$

in particular, for $(m,q) = 1$,

$$\lambda_f(m)\rho_f(1) = \sqrt{m}\rho_f(m). \tag{2.8}$$

Moreover, these relations hold for all m, n if f is a newform.

We will also need the following lower bounds for any L^2-normalized newform f in either $\mathcal{B}_k(q,\chi)$ or $\mathcal{B}(q,\chi)$:

$$|\rho_f(1)|^2 \gg_\varepsilon \begin{cases} (4\pi)^{k-1}((k-1)!q)^{-1}(kq)^{-\varepsilon}, & \text{for} \quad f \in \mathcal{B}_k(q,\chi), \\ \cosh(\pi t_f) q^{-1}(1+|t_f|)^{-\kappa}(q+|t_f|)^{-\varepsilon}, & \text{for} \quad f \in \mathcal{B}(q,\chi), \end{cases} \tag{2.9}$$

cf. [DFI5, (6.22)–(6.23), (7.15)–(7.16)] and [HM, (31)].

2.1.4. *The Kuznetsov formula.* Let $\phi : [0,\infty) \to \mathbb{C}$ be a smooth function satisfying $\phi(0) = \phi'(0) = 0$, $\phi^{(j)}(x) \ll_\varepsilon (1+x)^{-2-\varepsilon}$ for $0 \leqslant j \leqslant 3$. For $\kappa \in \{0,1\}$ let[4]

$$\begin{aligned} \dot{\phi}(k) &:= i^k \int_0^\infty J_{k-1}(x)\phi(x)\frac{dx}{x}, \\ \tilde{\phi}(t) &:= \frac{it^\kappa}{2\sinh(\pi t)} \int_0^\infty \left(J_{2it}(x) - (-1)^\kappa J_{-2it}(x)\right)\phi(x)\frac{dx}{x}, \\ \check{\phi}(t) &:= \frac{2}{\pi}\cosh(\pi t) \int_0^\infty K_{2it}(x)\phi(x)\frac{dx}{x} \end{aligned} \tag{2.10}$$

[4]For $\kappa \geqslant 2$ the definition of $\tilde{\phi}$ is different, and t^κ has to be replaced by a certain polynomial of degree κ; see [Pr, (84)] for the general formula.

be Bessel transforms. Then for positive integers m, n we have the following versions of the trace formula:

$$\begin{aligned}\sum_{q|c}\frac{1}{c}S_\chi(m,n,c)\phi\left(\frac{4\pi\sqrt{mn}}{c}\right) &= \sum_{\substack{k\equiv\kappa\,(2),\ k>\kappa\\ f\in\mathcal{B}_k(q,\chi)}}\sum \dot\phi(k)\frac{(k-1)!\sqrt{mn}}{\pi(4\pi)^{k-1}}\overline{\rho_f(m)}\rho_f(n)\\ &+\sum_{f\in\mathcal{B}(q,\chi)}\tilde\phi(t_f)\frac{4\pi\sqrt{mn}}{\cosh(\pi t_f)}\overline{\rho_f(m)}\rho_f(n)\\ &+\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\tilde\phi(t)\frac{\sqrt{mn}}{\cosh(\pi t)}\overline{\rho_f\,(m,t)}\rho_f\,(n,t)\;dt,\end{aligned}\tag{2.11}$$

and

$$\begin{aligned}\sum_{q|c}\frac{1}{c}S_\chi(m,-n,c)\phi\left(\frac{4\pi\sqrt{mn}}{c}\right) &= \sum_{f\in\mathcal{B}(q,\chi)}\check\phi(t_f)\frac{4\pi\sqrt{mn}}{\cosh(\pi t_f)}\overline{\rho_f(m)}\rho_f(-n)\\ &+\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\check\phi(t)\frac{\sqrt{mn}}{\cosh(\pi t)}\overline{\rho_f\,(m,t)}\rho_f\,(-n,t)\;dt,\end{aligned}\tag{2.12}$$

where the right-hand side runs over the spectrum of the Laplacian of weight $\kappa\in\{0,1\}$ in (2.11) and of weight $\kappa=0$ in (2.12), acting on forms of level q and character χ; alternatively, the Eisenstein contribution

$$\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\check\phi(t)\frac{\sqrt{mn}}{\cosh(\pi t)}\overline{\rho_f\,(m,t)}\rho_f\,(n,t)\;dt$$

can be replaced by the more usual sum

$$\sum_{\mathfrak{a}}\int_{-\infty}^{\infty}\check\phi(t)\frac{\sqrt{mn}}{\cosh(\pi t)}\overline{\rho_{\mathfrak{a}}\,(m,t)}\rho_{\mathfrak{a}}\,(n,t)\;dt,$$

where $\{\mathfrak{a}\}$ denotes the set of cusps for $\Gamma_0(q)$ which are singular with respect to χ; for a proof of the latter, see [Iw, Theorems 9.4 and 9.8][5] and [Pr]; the proof of these formulae with the Eisenstein parameters (2.1) is identical.

The holomorphic counterpart of (2.11)–(2.12) is Petersson's trace formula (cf. [Iw, Theorem 9.6] and [IK, Proposition 14.5])

$$\delta_{mn}+2\pi i^{-k}\sum_{q|c}\frac{1}{c}S_\chi(m,n,c)J_{k-1}\left(\frac{4\pi\sqrt{mn}}{c}\right)=\frac{(k-2)!\sqrt{mn}}{(4\pi)^{k-1}}\sum_{f\in\mathcal{B}_k(q,\chi)}\overline{\rho_f(m)}\rho_f(n).\tag{2.13}$$

[5]Note that in [Iw] a few misprints occur: equation (9.15) should have the normalization factor $\frac{2}{\pi}$ instead of $\frac{4}{\pi}$, and in equation (B.49) a factor 4 is missing.

Using the same notation[6] as above, the large sieve inequalities [DI, Theorem 2] state, for forms of level q and trivial nebentypus χ_0, that

$$\left.\begin{aligned}
&\sum_{\substack{k\equiv 0\,(2)\\ 2\leqslant k\leqslant T}} \frac{(k-1)!}{(4\pi)^{k-1}} \sum_{f\in\mathcal{B}_k(q,\chi_0)} \left|\sum_{M\leqslant m<2M} b_m\sqrt{m}\rho_f(m)\right|^2\\
&\sum_{\substack{f\in\mathcal{B}(q,\chi_0)\\ |t_f|\leqslant T}} \frac{1}{\cosh(\pi t_f)} \left|\sum_{M\leqslant m<2M} b_m\sqrt{m}\rho_f(m)\right|^2\\
&\sum_{\mathfrak{a}} \int_{-T}^{T} \frac{1}{\cosh(\pi t)} \left|\sum_{M\leqslant m<2M} b_m\sqrt{m}\rho_{\mathfrak{a}}(m,t)\right|^2 dt
\end{aligned}\right\} \ll_\varepsilon \left(T^2+\frac{M^{1+\varepsilon}}{q}\right)\sum_{M\leqslant m<2M}|b_m|^2, \tag{2.14}$$

where $M, T \geqslant 1$ and (b_m) is an arbitrary sequence of complex numbers (notice that here we have considered the classical family of Eisenstein series indexed by the cusps of $\Gamma_0(q)$). For individual m, we have the estimates

$$\begin{aligned}
&\sum_{\substack{k\equiv 0\,(2)\\ 2\leqslant k\leqslant T}} \frac{(k-1)!}{(4\pi)^{k-1}} \sum_{f\in\mathcal{B}_k(q,\chi)} m|\rho_f(m)|^2 \ll_\varepsilon (qTm)^\varepsilon T^2,\\
&\sum_{\substack{f\in\mathcal{B}(q,\chi)\\ |t_f|\leqslant T}} \frac{m|\rho_f(m)|^2}{\cosh(\pi t_f)} \ll_\varepsilon (qTm)^\varepsilon T^2 m^{2\theta},
\end{aligned} \tag{2.15}$$

see [HM, (35) and (37)].

2.2. **Special functions.** Special functions, in particular Bessel functions, will make an appearance at several places in this paper. If $\mathcal{B}_\nu$ denotes any of the Bessel functions J_ν, Y_ν, K_ν, then

$$(t^\nu \mathcal{B}_\nu(t))' = \pm t^\nu \mathcal{B}_{\nu-1}(t),$$

so that successive integration by parts yields

$$\int_0^\infty F(x)\mathcal{B}_0(\alpha\sqrt{x})\,dx = \left(\pm\frac{2}{\alpha}\right)^j \int_0^\infty x^{j/2}F^{(j)}(x)\mathcal{B}_j(\alpha\sqrt{x})\,dx \tag{2.16}$$

for $\alpha > 0$, $j \in \mathbb{N}_0$, and $F \in \mathcal{C}_0^\infty((0,\infty))$.

Lemma 1. *a) Let $\phi(x)$ be a smooth function supported on $x \asymp X$ such that $\phi^{(j)}(x) \ll_j X^{-j}$ for all $j \in \mathbb{N}_0$. For $t \in \mathbb{R}$ we have*

$$\dot\phi(t),\ \tilde\phi(t),\ \check\phi(t) \ll_C \frac{1+|\log X|}{1+X}\left(\frac{1+X}{1+|t|}\right)^C$$

for any constant $C \geqslant 0$. Here the Bessel transform $\tilde\phi$ is taken with respect to $\kappa = 0$.

b) Let $\phi(x)$ be a smooth function supported on $x \asymp X$ such that $\phi^{(j)}(x) \ll_j (X/Z)^{-j}$ for all $j \in \mathbb{N}_0$. For $t \in (-i/4, i/4)$ we have

$$\tilde\phi(t),\ \check\phi(t) \ll \frac{1+(X/Z)^{-2|\Im\mathfrak{m}\, t|}}{1+X/Z}.$$

Here the Bessel transform $\tilde\phi$ is taken with respect to $\kappa = 0$.

[6] Note the different normalization of the Fourier coefficients of Eisenstein series in [DI].

c) Assume that $\phi(x) = e^{iax}\psi(x)$ for some constant a and some smooth function $\psi(x)$ supported on $x \asymp X$ such that $\psi^{(j)}(x) \ll X^{-j}$ for all $j \in \mathbb{N}_0$. Assume $aX \geqslant 1$, $t \in \mathbb{R}$, and assume $t \in \mathbb{N}$ in the case of $\dot\phi$. Then

$$\dot\phi(t),\ \tilde\phi(t),\ \check\phi(t) \ll_{C,\varepsilon} \frac{1}{F^{1-\varepsilon}}\left(\frac{F}{1+|t|}\right)^C$$

for any $C \geqslant 0$, $\varepsilon > 0$ and some $F = F(X,a) < (a+1)X$.

Proof. Part a) is [DI, (7.1)–(7.2)] and [BHM, (2.14)]. Part b) is [BHM, (2.11)]. Part c) is [Ju, pp. 43–45]. □

For future reference we recast $\tilde\phi$, defined in (2.10), as follows: by [GR, 6.561.14] the Mellin transform of the Bessel kernel $k_t(x) := J_{2it}(x) - (-1)^\kappa J_{-2it}(x)$ equals

$$\begin{aligned}\widehat{k_t}(s) &= \int_0^\infty k_t(x)x^{s-1}dx \\ &= \frac{2^{s-1}}{\pi}\Gamma\left(\frac{s}{2}+it\right)\Gamma\left(\frac{s}{2}-it\right)\left\{\sin\left(\pi\left(\frac{s}{2}-it\right)\right) - (-1)^\kappa \sin\left(\pi\left(\frac{s}{2}+it\right)\right)\right\}.\end{aligned}$$

Let

$$\phi^*(u) := \widehat{\phi}(-1-2u)2^{1+2u}. \tag{2.17}$$

Then by Plancherel's formula

$$\begin{aligned}\tilde\phi(t) &= \frac{it^\kappa}{2\sinh(\pi t)}\frac{1}{2\pi i}\int_{(\sigma)}\phi^*(u)\widehat{k_t}(1+2u)2^{-2u}\,du \\ &= \frac{1}{\pi}\left\{\begin{array}{c}1\\ it\coth(\pi t)\end{array}\right\}\frac{1}{2\pi i}\int_{(\sigma)}\phi^*(u)\Gamma\left(\frac{1}{2}+u+it\right)\Gamma\left(\frac{1}{2}+u-it\right)\left\{\begin{array}{c}-\sin(\pi u)\\ +\cos(\pi u)\end{array}\right\}du,\end{aligned} \tag{2.18}$$

where $-\frac12 + |\Im\mathfrak{m}\, t| < \sigma < 0$, and the upper (resp. lower) line refers to $\kappa = 0$ (resp. $\kappa = 1$).

For the proof of Theorem 2, the exact shape of the test function ϕ is in principle irrelevant. However, it will be convenient to construct it as a linear combination of the following explicit functions. For integers $0 \leqslant b < a$ with $a - b \equiv \kappa \pmod 2$ we take

$$\phi_{a,b}(x) := i^{b-a}J_a(x)x^{-b}. \tag{2.19}$$

In order to satisfy the decay conditions for Kuznetsov's trace formula, we assume $b \geqslant 2$. Then by [GR, 6.561.14] we obtain

$$\phi^*_{a,b}(u) = i^{b-a}2^{-b-1}\frac{\Gamma\left((a-b-1-2u)/2\right)}{\Gamma\left((3+a+b+2u)/2\right)} \ll_{a,b} (1+|\Im\mathfrak{m}\, u|)^{-b-2-2\Re\mathfrak{e}\, u}, \quad |\Re\mathfrak{e}\, u| \leqslant \frac{a-b-2}{2}, \tag{2.20}$$

while using [GR, 6.574.2] it is straightforward to verify that

$$\begin{aligned}\dot\phi_{a,b}(k) &= \frac{b!}{2^{b+1}\pi}\prod_{j=0}^b\left\{\left(\frac{(1-k)i}{2}\right)^2 + \left(\frac{a+b}{2}-j\right)^2\right\}^{-1} \asymp_{a,b} \pm k^{-2b-2}, \\ \tilde\phi_{a,b}(t) &= \frac{b!}{2^{b+1}\pi}\left\{\begin{array}{c}1\\ t\coth(\pi t)\end{array}\right\}\prod_{j=0}^b\left\{t^2 + \left(\frac{a+b}{2}-j\right)^2\right\}^{-1} \asymp_{a,b} (1+|t|)^{\kappa-2b-2}.\end{aligned} \tag{2.21}$$

In particular,

$$\begin{aligned}&\dot\phi_{a,b}(k) > 0 \quad \text{for} \quad 2 \leqslant k \leqslant a-b, \\ &\tilde\phi_{a,b}(t) > 0 \quad \text{for} \quad \text{all possible spectral parameters } t,\end{aligned} \tag{2.22}$$

since $|\mathfrak{Im}\, t| < \frac{1}{2}$ when $\kappa = 0$, and $t \in \mathbb{R}$ when $\kappa = 1$. Moreover, we see that for any fixed a, b as above and any even polynomial $\alpha \in \mathbb{C}[T]$ of degree $2d \leqslant 2b-4$ there is a linear combination

$$\phi(x) = \sum_{\nu=0}^{d} \beta_\nu \phi_{a-\nu,b-\nu}(x) \tag{2.23}$$

with β_ν depending on a, b and the coefficients of α such that

$$\dot{\phi}(k) = \dot{\phi}_{a,b}(k)\alpha\left(\tfrac{(1-k)i}{2}\right) \qquad \text{and} \qquad \tilde{\phi}(t) = \tilde{\phi}_{a,b}(t)\alpha(t). \tag{2.24}$$

2.3. **Divisor sums.** Let τ be the divisor function. Exponential sums involving the divisor function can be handled by Voronoi summation. Let

$$L_w(x) := \log x + 2\gamma - 2\log w, \tag{2.25}$$

where γ is Euler's constant, and let

$$\mathcal{J}^-(x) := -2\pi Y_0(4\pi x), \qquad \mathcal{J}^+(x) := 4K_0(4\pi x)$$

with the usual Bessel functions. For later purposes we write $\mathcal{J}^\pm$ as inverse Mellin transforms using [GR, 17.43.17, 17.43.18] or [KMV, (36)]:

$$\begin{aligned} \mathcal{J}^+(\sqrt{x}) &= \frac{2}{2\pi i}\int_{(1)} (2\pi)^{-2u}\Gamma(u)^2 x^{-u}\, du, \\ \mathcal{J}^-(\sqrt{x}) &= \frac{2}{2\pi i}\int_{(*)} (2\pi)^{-2u}\Gamma(u)^2 x^{-u}\cos(\pi u)\, du, \end{aligned} \tag{2.26}$$

where $(*)$ is the path $\mathfrak{Re}\, u = -1$ except when $|\mathfrak{Im}\, u| < 1$ where it curves to hit the real axis at $u > 0$. Let $(d,c) = 1$ and let $F \in \mathcal{C}_0^\infty((0,\infty))$, then

$$\begin{aligned} \sum_{m=1}^{\infty} \tau(m) e\left(\frac{dm}{c}\right) F(m) =& \frac{1}{c}\int_0^\infty L_c(y)F(y)\, dy \\ &+ \frac{1}{c}\sum_{\pm}\sum_{m=1}^{\infty} \tau(m)\, e\left(\frac{\pm\bar{d}m}{c}\right)\int_0^\infty \mathcal{J}^\pm\left(\frac{\sqrt{my}}{c}\right) F(y)\, dy. \end{aligned} \tag{2.27}$$

In order to evaluate additive divisor sums, we use the following method, cf. [Me, (2.1) and (2.4)]. Here and later in the proof, we will need smooth cut-off functions. Let henceforth ω denote a smooth function such that $\omega(x) = 1$ on $[0,1]$ and $\omega(x) = 0$ on $[2,\infty)$. Then we have

$$\left(1 - \omega\left(\frac{x}{\sqrt{Q}}\right)\right)\left(1 - \omega\left(\frac{y}{x\sqrt{Q}}\right)\right) = 0$$

for all $x, y, Q \geqslant 0$ such that $y \leqslant Q$. Therefore

$$\tau(n) = \sum_{\delta|n} \omega\left(\frac{\delta}{\sqrt{Q}}\right)\left(2 - \omega\left(\frac{n}{\delta\sqrt{Q}}\right)\right)$$

whenever $n \leqslant Q$. Let $g : [\frac{1}{2}, Q] \times [\frac{1}{2}, M] \to \mathbb{C}$ be a smooth function. Then

$$\begin{aligned} &\sum_{an\pm m=h} \tau(n)\tau(m)g(n,m) = \sum_{n=1}^{\infty} \tau(n)\tau(\pm(h-an))g(n,\pm(h-an)) \\ &= \sum_{\delta=1}^{\infty} \omega\left(\frac{\delta}{\sqrt{Q}}\right)\sum_{\delta|n} \tau(\pm(h-an))g(n,\pm(h-an))\left(2 - \omega\left(\frac{n}{\delta\sqrt{Q}}\right)\right) \\ &= \sum_{\delta=1}^{\infty} \omega\left(\frac{\delta}{\sqrt{Q}}\right)\sum_{m \equiv \pm h\ (a\delta)} \tau(m) g\left(\frac{h \mp m}{a}, m\right)\left(2 - \omega\left(\frac{h\mp m}{a\delta\sqrt{Q}}\right)\right). \end{aligned}$$

Using additive characters and Voronoi summation (2.27), we get

$$\sum_{m\equiv\mu\,(c)}\tau(m)F(m)=\frac{1}{c}\sum_{w|c}\frac{r_w(\mu)}{w}\int_0^\infty L_w(y)F(y)\,dy$$
$$+\sum_{\pm}\frac{1}{c}\sum_{w|c}\frac{1}{w}\sum_{m=1}^\infty\tau(m)S(-\mu,\pm m;w)\int_0^\infty\mathcal{J}^\pm\left(\frac{\sqrt{my}}{w}\right)F(y)\,dy$$

for any compactly supported smooth function F, so that

(2.28)

$$\sum_{an\pm m=h}\tau(n)\tau(m)g(n,m)=\sum_{w=1}^\infty\frac{(a,w)r_w(h)}{w^2}\int_0^\infty L_w(\pm(h-ax))K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx$$
$$+\sum_{w=1}^\infty\frac{(a,w)}{w^2}\sum_{n=1}^\infty\tau(n)S\left(\mp h,n;w\right)\int_0^\infty\mathcal{J}^+\left(\frac{\sqrt{n(\pm(h-ax))}}{w}\right)K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx$$
$$+\sum_{w=1}^\infty\frac{(a,w)}{w^2}\sum_{n=1}^\infty\tau(n)S\left(\mp h,-n;w\right)\int_0^\infty\mathcal{J}^-\left(\frac{\sqrt{n(\pm(h-ax))}}{w}\right)K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx,$$

where

$$r_w(h):=S(h,0;w)=\sum_{d|(h,w)}d\mu(w/d)\tag{2.29}$$

is the Ramanujan sum and

$$K_{r,w}(x):=\sum_{\delta=1}^\infty\frac{1}{\delta}\omega\left(\frac{w\delta}{r\sqrt{Q}}\right)\left(2-\omega\left(\frac{rx}{\delta w\sqrt{Q}}\right)\right).$$

For future reference we state some properties of $K_{r,w}(x)$. A straightforward calculation shows

$$x^iw^j\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial w^j}K_{r,w}(x)\ll_{i,j}\log Q\tag{2.30}$$

for any $i,j\geqslant 0$, and clearly

$$K_{r,w}(x)=0\quad\text{if } w\geqslant 2r\sqrt{Q}.\tag{2.31}$$

3. Approximate functional equation and amplification

Let $f=f_0$ be a primitive (holomorphic or Maaß) cusp form having L^2-norm 1, for which we want to prove Theorem 2. Let t_0 denote its spectral parameter as defined in (1.1). For $\mathfrak{Re}\,s>1$ the L-function of f_0 is defined as a Dirichlet series in the Hecke eigenvalues of f_0

$$L(f_0,s):=\sum_{n=1}^\infty\lambda_{f_0}(n)n^{-s}.$$

The completed L-function is given by

$$\Lambda(f_0,s):=q^{s/2}L_\infty(f_0,s)L(f_0,s),\qquad L_\infty(f_0,s):=\pi^{-s}\Gamma\left(\frac{s+\mu_1}{2}\right)\Gamma\left(\frac{s+\mu_2}{2}\right),$$

where

$$\mu_1,\ \mu_2:=\begin{cases}it_0, & -it_0 & \text{when } f_0 \text{ is an even Maaß form of even weight;}\\ it_0, & -it_0+1 & \text{when } f_0 \text{ is an even Maaß form of odd weight;}\\ it_0+1, & -it_0+1 & \text{when } f_0 \text{ is an odd Maaß form of even weight;}\\ it_0+1, & -it_0 & \text{when } f_0 \text{ is an odd Maaß form of odd weight;}\\ it_0, & it_0+1 & \text{when } f_0 \text{ is a holomorphic form.}\end{cases}$$

Observe that (2.5) implies

$$\Re\mathfrak{e}\,\mu_1,\ \Re\mathfrak{e}\,\mu_2\ \geqslant\ -\frac{7}{64}. \tag{3.1}$$

The completed L-function is entire and satisfies the functional equation [DFI5, (8.11)–(8.13), (8.17)–(8.19)]

$$\Lambda(f_0,s)=\omega\overline{\Lambda}(f_0,1-\overline{s}) \tag{3.2}$$

for some constant $\omega=\omega(f_0)$ of modulus 1. Relation (2.2) shows that

$$L(f_0,s)^2=L(2s,\chi)\sum_{n=1}^{\infty}\tau(n)\lambda_{f_0}(n)n^{-s},\qquad \Re\mathfrak{e}\,s>1. \tag{3.3}$$

Let us fix a point s on the critical line $\Re\mathfrak{e}\,s=\frac{1}{2}$ for which we want to prove Theorem 2. The above Dirichlet series no longer converges (absolutely) for s but a similar formula holds which is traditionally called an approximate functional equation. In order to achieve polynomial dependence in the spectral parameter t_0 we will closely follow the argument in [Ha] specified for the shifted L-function $u\mapsto L\left(f_0,s-\frac{1}{2}+u\right)$. We define the analytic conductor [Ha, (2.4) and Remark 2.7]

$$C=C(f_0,s):=\frac{q}{(2\pi)^2}|s+\mu_1||s+\mu_2| \tag{3.4}$$

and the auxiliary function [Ha, (1) in Erratum]

$$F(f_0,s;u):=\frac{1}{2}C^{-u/2}q^u\frac{L_\infty(f_0,s+u)\overline{L_\infty}(f_0,s)}{\overline{L_\infty}(f_0,s-\overline{u})L_\infty(f_0,s)}+\frac{1}{2}C^{u/2}.$$

By (3.1) this function is holomorphic in $\Re\mathfrak{e}\,u>-\frac{1}{4}$ (say) and satisfies the bound [Ha, (2) in Erratum]

$$C^{-u/2}F(f_0,s;u)-\frac{1}{2}\ll_\sigma(1+|u|)^{2\Re\mathfrak{e}\,u},\qquad -\frac{1}{4}<\Re\mathfrak{e}\,u\leqslant\sigma \tag{3.5}$$

with an implied constant independent of s and f_0. In addition, we have $F(f_0,s;0)=1$, and from the functional equation (3.2) we can deduce [Ha, (3.3)]

$$F(f_0,s;u)L(f_0,s+u)=\omega\lambda\overline{F}(f_0,s;-\overline{u})\overline{L}(f_0,s-\overline{u}),\qquad \lambda:=\frac{\overline{L_\infty}(f_0,s)}{L_\infty(f_0,s)}.$$

In particular,

$$\eta=\eta(f_0,s):=(\omega\lambda)^2$$

is of modulus 1 and with the notation

$$G^+(u):=F(f_0,s;\tfrac{1}{2}-s+u)^2,\qquad G^-(u):=\overline{F}(f_0,s;\tfrac{1}{2}-s+\overline{u})^2$$

we obtain the functional equation

$$G^+(u)L(f_0,\tfrac{1}{2}+u)^2=\eta G^-(-u)\overline{L}(f_0,\tfrac{1}{2}-\overline{u})^2. \tag{3.6}$$

Observe that (3.5) implies, for $0<\varepsilon\leqslant\Re\mathfrak{e}\,u\leqslant\sigma$,

$$G^\pm(u)\ll_{\varepsilon,\sigma}C^{\Re\mathfrak{e}\,u}\left(1+|\mathfrak{Im}\,u\mp\mathfrak{Im}\,s|\right)^{4\Re\mathfrak{e}\,u}. \tag{3.7}$$

We fix an arbitrary entire function $P(u)$ which decays fast in vertical strips and satisfies $P(0)=1$ as well as $P(u)=P(-u)=\overline{P(\overline{u})}$. The role of this factor is to make the dependence on s in Theorem 2 polynomial. We introduce another even function in order to create zeros that avoid the matching, as discussed in Section 1.2:

$$Q(u,t):=\left(u^2-(\tfrac{1}{2}-it)^2\right)^2\left(u^2-(\tfrac{1}{2}+it)^2\right)^2=:\sum_{\nu=0}^{4}\alpha_\nu(t)u^{2\nu} \tag{3.8}$$

for suitable real even polynomials $\alpha_\nu\in\mathbb{R}[T]$. Note that

$$Q\left(u,i\left(\tfrac{1}{2}-u\right)\right)=Q^{(1,0)}\left(u,i\left(\tfrac{1}{2}-u\right)\right)=0. \tag{3.9}$$

Now we apply the usual contour shift technique to the integral

$$\frac{1}{2\pi i}\int_{(1)} L\left(f_0, \tfrac{1}{2}+u\right)^2 G^+(u)P(u+\tfrac{1}{2}-s)\frac{Q(u,t_0)}{Q(s-\frac{1}{2},t_0)}\cdot\frac{du}{u+\frac{1}{2}-s}.$$

In combination with (3.3) and (3.6) we obtain

$$L(f_0,s)^2 = \sum_{n=1}^{\infty}\frac{\tau(n)\lambda_{f_0}(n)V_{t_0}^+(n/q)}{n^{1/2}} + \eta\sum_{n=1}^{\infty}\frac{\tau(n)\overline{\lambda_{f_0}(n)}V_{t_0}^-(n/q)}{n^{1/2}}, \tag{3.10}$$

where we define $V_t^{\pm}$ for any spectral parameter t through its Mellin transform

$$\begin{aligned}\widehat{V}_t^+(u) &:= \widehat{W}^+(u)Q(u,t) := q^{-u}G^+(u)L(1+2u,\chi)\frac{P(u+\frac{1}{2}-s)}{u+\frac{1}{2}-s}\cdot\frac{Q(u,t)}{Q\left(s-\frac{1}{2},t_0\right)},\\ \widehat{V}_t^-(u) &:= \widehat{W}^-(u)Q(u,t) := q^{-u}G^-(u)L(1+2u,\overline{\chi})\frac{P(u+s-\frac{1}{2})}{u+s-\frac{1}{2}}\cdot\frac{Q(u,t)}{Q\left(s-\frac{1}{2},t_0\right)}.\end{aligned} \tag{3.11}$$

Here we have suppressed the notational dependence of $\widehat{V}_t^{\pm}$ and $\widehat{W}^{\pm}$ on s and t_0 as these parameters are kept fixed in the rest of the paper. Since $Q\left(s-\frac{1}{2},t_0\right)$ is real for $\mathfrak{Re}\, s = \frac{1}{2}$ and the spectral parameter t_0, we have

$$\begin{aligned}\widehat{W}^-(u) &= \overline{\widehat{W}^+}(\overline{u}) \qquad \text{and} \qquad & \widehat{V}_t^-(u) &= \overline{\widehat{V}_t^+}(\overline{u})\\ W^-(x) &= \overline{W^+}(x) & V_t^-(x) &= \overline{V_t^+}(x).\end{aligned} \tag{3.12}$$

We can therefore drop the superscripts and write

$$W := W^+ \qquad \text{and} \qquad V_t := V_t^+.$$

Note that by (3.11) and (3.8),

$$V_t(x) = \sum_{\nu=0}^{4}\alpha_\nu(t)\left(x\frac{\partial}{\partial x}\right)^{2\nu} W(x). \tag{3.13}$$

By (3.7), (3.4), (3.8), (1.1), (2.5), it follows, for $0<\varepsilon\leqslant\mathfrak{Re}\, u\leqslant\sigma$ and for any $A>0$,

$$\widehat{W}(u) \ll_{\varepsilon,\sigma,A} \left(|s|+|t_0|\right)^{2\mathfrak{Re}\, u}\left(1+|\mathfrak{Im}\, u \mp \mathfrak{Im}\, s|\right)^{-A}.$$

Therefore $\widehat{W}$ is rapidly decaying on vertical lines and inverse Mellin transformation shows

$$x^i\frac{\partial^i}{\partial x^i}W(x) \ll_{\varepsilon,B,i} |s|^{i+1}(|s|+|t_0|)^{2\varepsilon}x^{-\varepsilon}\left(1+\frac{x}{(|s|+|t_0|)^2}\right)^{-B}, \qquad B,i\in\mathbb{N}_0. \tag{3.14}$$

With these auxiliary functions we introduce the following family of "fake" L-functions for any cusp form f either in $\mathcal{B}_k(q,\chi)$ or in $\mathcal{B}(q,\chi)$ and for any Eisenstein series E_{χ_1,χ_2}:

$$\begin{aligned}\Sigma(f\otimes E,s) &:= \sum_{n=1}^{\infty}\frac{\tau(n)\sqrt{n}\rho_f(n)V_{t_f}(n/q)}{n^{1/2}},\\ \Sigma(E_{\chi_1,\chi_2,f}(\cdot,\tfrac{1}{2}+it)\otimes E,s) &:= \sum_{n=1}^{\infty}\frac{\tau(n)\sqrt{n}\rho_f(n,t)V_t(n/q)}{n^{1/2}}.\end{aligned} \tag{3.15}$$

With this notation (3.10) reads for $f=f_0$ (cf. (2.8) and (3.12))

$$\rho_{f_0}(1)L(f_0,s)^2 = \Sigma(f_0\otimes E,s) + \eta(f_0,s)\overline{\Sigma(f_0\otimes E,s)}. \tag{3.16}$$

In order to apply the trace formula, we wanted an approximate functional equation that is "as independent of t_0 as possible"; now the information on the spectral parameter is all encoded in the polynomial $Q(u,t)$. In [DFI5], however, the weight function was the same for all the f's which made the rest of the proof more complicated.

For every given $\ell \in \mathbb{N}$ satisfying $(\ell, q) = 1$ and $\ell \leqslant q$ let us define with the notation of Section 2.1

$$\mathcal{Q}_k^{\mathrm{holo}}(\ell) := \frac{i^k(k-2)!}{2\pi(4\pi)^{k-1}} \sum_{f \in \mathcal{B}_k(q,\chi)} \lambda_f(\ell) \big|\Sigma(f \otimes E, s)\big|^2,$$

$$\begin{aligned} \mathcal{Q}(\ell) := & \sum_{\substack{k \equiv \kappa\,(2) \\ k > \kappa}} \dot{\phi}_0(k) 2(k-1) i^{-k} \mathcal{Q}_k^{\mathrm{holo}}(\ell) + \sum_{f \in \mathcal{B}(q,\chi)} \tilde{\phi}_0(t_f) \frac{4\pi}{\cosh(\pi t_f)} \lambda_f(\ell) \big|\Sigma(f \otimes E, s)\big|^2 \\ & + \sum_{\substack{\chi_1\chi_2 = \chi \\ f \in \mathcal{B}(\chi_1,\chi_2)}} \sum \int_{-\infty}^{\infty} \tilde{\phi}_0(t) \frac{1}{\cosh(\pi t)} \lambda_{\chi_1,\chi_2}(\ell, t) \big|\Sigma(E_{\chi_1,\chi_2,f}(\cdot, \tfrac{1}{2} + it) \otimes E, s)\big|^2 \, dt, \end{aligned} \tag{3.17}$$

where

$$\phi_0(x) := \phi_{A,10}(x) = i^{10-A} J_A(x) x^{-10}, \tag{3.18}$$

cf. (2.19) for some very large A of parity κ.

Remark 3.1. Let us explain the reason of our choice for the construction of $V_t^{\pm}$. Suppose for simplicity that q is prime (hence χ being non-trivial is primitive). In that case there are two Eisenstein series $E_{\chi,1}(z, f, \frac{1}{2} + it)$ and $E_{1,\chi}(z, f, \frac{1}{2} + it)$ (which are the Eisenstein series associated to the cusps $\mathfrak{a} = 0, \infty$). Their contribution to the above sum equals

$$\int_{-\infty}^{\infty} \tilde{\phi}_0(t) \frac{|\rho(1,t)|^2}{\cosh(\pi t)} \lambda_{\chi,1}(\ell, t) \left| \frac{1}{2\pi i} \int_{(1)} \frac{L^2(u + \frac{1}{2} - it, \chi) \zeta^2(u + \frac{1}{2} + it)}{L(2u+1, \chi)} \widehat{V}_t(u) \, q^u \, du \right|^2 dt. \tag{3.19}$$

The main contribution comes from the double pole of the inner integrand, and we designed V_t such that it kills this pole. This is reflected by the vanishing of (5.10) below.

We shall show

$$k^{-18} \left|\mathcal{Q}_k^{\mathrm{holo}}(\ell)\right| + |\mathcal{Q}(\ell)| \ll_{s,t_0,\varepsilon} q^{\varepsilon} \left(\ell^{c_1} q^{-c_2} + \ell^{-1/2}\right) \tag{3.20}$$

for certain positive absolute constants c_1 and c_2, uniformly in $k \geqslant A - 10$, and with polynomial dependence on s and t_0. This implies Theorem 2: Let us choose the standard amplifier

$$x(\ell) := \begin{cases} \lambda(p)\overline{\chi}(p) & \text{if} \quad \ell = p, \quad p \nmid q, \quad \frac{1}{2}\sqrt{L} < p \leqslant \sqrt{L}; \\ -\overline{\chi}(p) & \text{if} \quad \ell = p^2, \quad p \nmid q, \quad \frac{1}{2}\sqrt{L} < p \leqslant \sqrt{L}; \\ 0 & \text{else;} \end{cases}$$

for some parameter $\log L \asymp \log q$ to be chosen in a minute. Using (2.2) with $n = m = p$, we see

$$\sum_{\ell} x(\ell)\lambda(\ell) = \sum_{\substack{p \nmid q \\ \frac{1}{2}\sqrt{L} < p \leqslant \sqrt{L}}} 1 \gg L^{1/2-\varepsilon}.$$

Therefore, by (3.16), (2.9), (3.18), (2.21)–(2.22), we obtain

$$\begin{aligned}\frac{L}{q^{1+\varepsilon}}\big|L(f_0,s)\big|^4 \ll_{t_0,\varepsilon}\ & \sum_{\substack{k\equiv\kappa\,(2),\ k>\kappa\\ f\in\mathcal{B}_k(q,\chi)}}\sum \big|\dot\phi_0(k)\big|\frac{(k-1)!}{\pi(4\pi)^{k-1}}\left|\sum_\ell x(\ell)\lambda_f(\ell)\right|^2\big|\Sigma(f\otimes E,s)\big|^2\\ &+\sum_{f\in\mathcal{B}(q,\chi)}\tilde\phi_0(t_f)\frac{4\pi}{\cosh(\pi t_f)}\left|\sum_\ell x(\ell)\lambda_f(\ell)\right|^2\big|\Sigma(f\otimes E,s)\big|^2\\ &+\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\tilde\phi_0(t)\frac{1}{\pi\cosh(\pi t)}\left|\sum_\ell x(\ell)\lambda_{\chi_1,\chi_2}(\ell,t)\right|^2\big|\Sigma(E_{\chi_1,\chi_2,f}(\cdot,\tfrac12+it)\otimes E,s)\big|^2\,dt,\end{aligned}$$

so that by (2.3), (2.22) and (3.17) we obtain

$$\begin{aligned}&\frac{L}{q^{1+\varepsilon}}\big|L(f_0,s)\big|^4 \ll_{t_0,\varepsilon}\\ &\sum_{\ell_1,\ell_2}\big|x(\ell_1)x(\ell_2)\big|\sum_{d\mid(l_1,l_2)}\left\{\left|\mathcal{Q}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|+\sum_{\substack{k\equiv\kappa\,(2)\\ k\geqslant A-10}}4k\big|\dot\phi_0(k)\big|\left|\mathcal{Q}_k^{\mathrm{holo}}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|\right\}.\end{aligned}\tag{3.21}$$

Substituting (3.20) (note that the k-sum converges by (2.21)) and changing the order of summation, this is

$$\begin{aligned}&\ll_{s,t_0,\varepsilon} q^{\varepsilon}\left\{q^{-c_2}\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{c_1}|x(d\ell_1)x(d\ell_2)|+\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{-1/2}|x(d\ell_1)x(d\ell_2)|\right\}\\ &\ll_{s,t_0,\varepsilon} q^{\varepsilon}\left(L^{2c_1+1/2}q^{-c_2}+1\right)\sum_\ell\tau(\ell)|x(\ell)|^2,\end{aligned}$$

where we used Cauchy–Schwarz twice. By (2.6), we obtain from the last two displays

$$|L(f_0,s)|^4 \ll_{s,t_0,\varepsilon} q^{1+\varepsilon}\left(L^{2c_1}q^{-c_2}+L^{-1/2}\right).$$

Choosing

$$L := q^{c_2/(2c_1+1/2)},\tag{3.22}$$

this gives Theorem 2 with

$$L(f_0,s) \ll_{s,t_0,\varepsilon} q^{\frac14-\frac{c_2}{4(4c_1+1)}+\varepsilon}.\tag{3.23}$$

It remains to show (3.20) and calculate the constants c_1 and c_2. This will be done in the next three sections.

4. Applying Kuznetsov and Voronoi summation formulae

As a first step we substitute (3.15) into the definition (3.17) of $\mathcal{Q}_k^{\mathrm{holo}}(\ell)$ and $\mathcal{Q}(\ell)$. Then we apply (2.2) and the corresponding formula for the divisor function in order to remove the factors $\lambda_f(\ell)$

and $\lambda_{\chi_1,\chi_2}(t,\ell)$. Applying (2.7), this gives

$$k^{-18}\mathcal{Q}_k^{\mathrm{holo}}(\ell) = \sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{\sqrt{a}}\sum_{m,n}\frac{\tau(m)\tau(n)}{(mn)^{1/2}} \times \overline{k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{m}{q}\right)}k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\frac{i^k(k-2)!\sqrt{maen}}{2\pi(4\pi)^{k-1}}\sum_{f\in\mathcal{B}_k(q,\chi)}\overline{\rho_f(m)}\rho_f(aen)$$

and

$$\begin{aligned}\mathcal{Q}(\ell) = \sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{\sqrt{a}}\sum_{m,n}\frac{\tau(m)\tau(n)}{(mn)^{1/2}}\\ \times\Bigg\{\sum_{f\in\mathcal{B}(q,\chi)}\tilde{\phi}_0(t_f)\overline{V_{t_f}\left(\frac{m}{q}\right)}V_{t_f}\left(\frac{adn}{q}\right)\frac{4\pi\sqrt{maen}}{\cosh(\pi t_f)}\overline{\rho_f(m)}\rho_f(aen)\\ +\sum_{\chi_1\chi_2=\chi}\sum_{f\in\mathcal{B}(\chi_1,\chi_2)}\int_{-\infty}^{\infty}\tilde{\phi}_0(t)\overline{V_t\left(\frac{m}{q}\right)}V_t\left(\frac{adn}{q}\right)\frac{\sqrt{maen}}{\cosh(\pi t)}\overline{\rho_f(m,t)}\rho_f(aen,t)\,dt\\ +\sum_{k\equiv\kappa\,(2),\ k>\kappa}\sum_{f\in\mathcal{B}_k(q,\chi)}\dot{\phi}_0(k)\overline{V_{\frac{(1-k)i}{2}}\left(\frac{m}{q}\right)}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\frac{(k-1)!\sqrt{maen}}{\pi(4\pi)^{k-1}}\overline{\rho_f(m)}\rho_f(aen)\Bigg\}.\end{aligned}$$

Substituting (3.13), we get something of the form

$$\begin{aligned}\mathcal{Q}(\ell) = \sum_{\nu,\xi=0}^{4}\dots\Bigg\{\sum_j\tilde{\phi}_0(t_f)\overline{\alpha_\nu(t_f)}\alpha_\xi(t_f)\overline{\left(x\frac{\partial}{\partial x}\right)^{2\nu}W\left(\frac{m}{q}\right)}\left(x\frac{\partial}{\partial x}\right)^{2\xi}W\left(\frac{adn}{q}\right)\dots\\ +\text{ Eisenstein contribution }+\text{ holomorphic contribution }\Bigg\}.\end{aligned}\tag{4.1}$$

Now we apply Kuznetsov's trace formula (2.11) for each term separately. Similarly, we apply Petersson's formula (2.13) for $\mathcal{Q}_k^{\mathrm{holo}}(\ell)$. In the latter case we obtain a diagonal term which can be estimated trivially using (3.13) and (3.14):

$$\frac{i^k}{2\pi}\sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{a\sqrt{e}}\sum_n\frac{\tau(aen)\tau(n)}{n}\overline{k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{aen}{q}\right)}k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\ll_\varepsilon q^\varepsilon\ell^{-1/2}.\tag{4.2}$$

Here and henceforth we suppress the dependence on s and t_0 and merely make sure that it is polynomial at most. In either case the off-diagonal term is a linear combination of terms of the form

$$\sum_{abe=\ell}\frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}}\sum_{q|c}\frac{1}{c}\sum_{m,n}\frac{\tau(m)\tau(n)}{(mn)^{1/2}}W_1\left(\frac{m}{q}\right)W_2\left(\frac{a^2bn}{q}\right)S_\chi(m,aen;c)\phi\left(\frac{4\pi\sqrt{aemn}}{c}\right).\tag{4.3}$$

where ϕ is J_{k-1} or a suitable ϕ as in (2.23)–(2.24) (with $a:=A$, $b:=10$, $\alpha:=\overline{\alpha_\nu}\alpha_\xi$ and $d:=8$), cf. (3.18). In particular, by (3.14), (2.23), (2.19)–(2.20),

$$W_{1,2}^{(i)}(x)\ll_{\varepsilon,B,i}x^{-i-\varepsilon}(1+x)^{-B},\quad \phi^{(i)}(x)\ll_{A,i}\left(\frac{x}{1+x}\right)^{A-10-i},\quad \phi^*(u)\ll(1+|\mathfrak{Im}\,u|)^{-2-2\mathfrak{Re}\,u}\tag{4.4}$$

for all i with some very large A, B and for all u in a wide vertical strip symmetric about the origin.

Let us now open the Kloosterman sum and apply Voronoi summation (2.27) to the m-variable. It is one of the main features of the Voronoi summation here that the twisted Kloosterman sum

becomes a Gauß sum. Let

$$G_\chi(h;c) := \sum_{\substack{d \pmod c \\ (d,c)=1}} \chi(d) e\left(\frac{hd}{c}\right)$$

denote the Gauß sum, then the term (4.3) decomposes into the sum of a "diagonal" first term

$$\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \sum_n \frac{\tau(n) G_{\overline{\chi}}(aen;c)}{n^{1/2}} W_2\left(\frac{a^2bn}{q}\right) \times \int_0^\infty L_c(y) W_1\left(\frac{y}{q}\right) \phi\left(\frac{4\pi\sqrt{aeny}}{c}\right) \frac{dy}{y^{1/2}}, \tag{4.5}$$

and of an "off-diagonal" second term given by

$$\sum_{\pm} \sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \sum_h G_{\overline{\chi}}(h;c) \sum_{aen \pm m = h} \tau(m)\tau(n) g^{\pm}(n,m;c), \tag{4.6}$$

where

$$g^{\pm}(n,m;c) := \frac{1}{n^{1/2}} W_2\left(\frac{a^2bn}{q}\right) \int_0^\infty \mathcal{J}^{\pm}\left(\frac{\sqrt{my}}{c}\right) W_1\left(\frac{y}{q}\right) \phi\left(\frac{4\pi\sqrt{aeny}}{c}\right) \frac{dy}{y^{1/2}} \tag{4.7}$$

for $c \geqslant q$.

Using the weak bound (cf. (5.5))

$$|G_{\overline{\chi}}(h;c)| \leqslant c^{1/2}(c,h)^{1/2},$$

the fact that $(\ell, q) = 1$ and also the inequalities (4.4) (cf. (4.9)), we obtain that (4.5) is bounded by

$$\ll_\varepsilon q^\varepsilon \sum_{\substack{q|c \\ c \leqslant \ell^{1/2} q^{1+\varepsilon}}} \frac{1}{c^2} \sum_{n \leqslant q^{1+\varepsilon}} \frac{c^{1/2}(c,\ell n)^{1/2}}{n^{1/2-\varepsilon}} q^{1/2+\varepsilon} \ll_\varepsilon q^{3\varepsilon - 1/2}, \tag{4.8}$$

As for the term (4.6), let us attach a smooth factor $\psi(m)$ to $g^{\pm}$ that is zero for $m \leqslant 1/2$ and 1 for $m \geqslant 3/4$. This does not affect the sum (4.6). We need this little technicality in order to apply (2.28) later. It is easy to see that $g^{\pm}(n,m;c)$ is negligible (i.e., $\ll q^{-C}$ for any constant $C > 0$) unless

$$\frac{q^{1-\varepsilon}}{ae} =: N^- \leqslant n \leqslant N^+ := \frac{q^{1+\varepsilon}}{a^2 b}, \qquad c \leqslant \frac{\sqrt{e} q^{1+\varepsilon}}{\sqrt{ab}}, \qquad m \leqslant aenq^\varepsilon. \tag{4.9}$$

The upper bound on n follows directly from (4.4) by choosing A and B large enough. By (4.4) we can also assume that $cq^{-\varepsilon} \leqslant \sqrt{aeny}$ and $y \leqslant q^{1+\varepsilon}$. Combining these inequalities, we obtain $c^2 q^{-3\varepsilon} \leqslant qaen$ which implies the lower bound on n and, in combination with the upper bound on n, it implies the upper bound on c as well. Finally, the upper bound on m follows from (2.16) by choosing a large j there. As a by-product, we can see that the integral in (4.7) is essentially supported on $[q^{1-\varepsilon}e^{-1}, q^{1+\varepsilon}]$, hence by applying a crude bound for the Bessel functions in that integral (e.g. [HM, Appendix]) we obtain

$$g^{\pm}(n,m;c) \ll_\varepsilon q^{1/2+\varepsilon} n^{-1/2} \quad \text{for} \quad n \leqslant q^{1+\varepsilon} \text{ and } c \geqslant q. \tag{4.10}$$

Let $\mathcal{S}(a,b,e,c;q)$ denote the weighted sum of shifted convolution sums

$$\mathcal{S}(a,b,e,c;q) := \sum_h G_{\overline{\chi}}(h;c) \sum_{\pm} \sum_{aen \pm m = h} \tau(m)\tau(n) g^{\pm}(n,m;c) \psi(m).$$

Thus (4.6) equals

$$\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \mathcal{S}(a,b,e,c;q). \tag{4.11}$$

Remark 4.1. Since we have assumed that χ is not trivial, $G_{\overline{\chi}}(0;c)=0$, hence in $\mathcal{S}(a,b,e,c;q)$ the h-sum varies over the $h\neq 0$. When χ is trivial, the degenerate contribution corresponding to $h=0$,

$$\mathcal{S}_0(a,b,e,c;q):=\varphi(c)\sum_{aen=m}\tau(m)\tau(n)g^-(n,m;c),$$

yields a main term which can be bounded by $\ll_\varepsilon q^\varepsilon \ell^{-1/2}$. We do not carry out this computation in this paper and rather refer to [KMV, Section 3.6].

Applying (2.28) with

$$Q:=N^+=\frac{q^{1+\varepsilon}}{a^2b} \tag{4.12}$$

to the innermost sum, $\mathcal{S}(a,b,e,c;q)$ splits into a main term

$$\begin{aligned}\mathcal{S}^M(a,b,e,c;q):=&\sum_{h\neq 0}G_{\overline{\chi}}(h;c)\sum_{\pm}\sum_{w=1}^{\infty}\frac{(ae,w)r_w(h)}{w^2}\\ &\times\int_0^\infty L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x,\pm(h-aex);c\right)\psi(\pm(h-aex))\,dx.\end{aligned} \tag{4.13}$$

and two error terms of the shape

$$\begin{aligned}\mathcal{S}^{E,\pm}(a,b,e,c;q):=&\sum_{h\neq 0}G_{\overline{\chi}}(h;c)\sum_{\pm}\sum_{w=1}^{\infty}\frac{(ae,w)}{w^2}\sum_{n=1}^{\infty}\tau(n)S\left(\mp h,\pm n;w\right)\\ &\times\int_0^\infty \mathcal{J}^{\pm}\left(\frac{\sqrt{n(\pm(h-aex))}}{w}\right)K_{(ae,w),w}(x)g^{\pm}(x,\pm(h-aex);c)\psi(\pm(h-aex))\,dx\end{aligned} \tag{4.14}$$

for various combinations of $\pm$. We postpone the estimation of (4.14) to Section 6, and start with the contribution of (4.13) to $\mathcal{S}(a,b,e,c;q)$. At this point, we need to remove the catalyst function $\psi(m)$ in (4.13) and define

$$\begin{aligned}\tilde{\mathcal{S}}^M(a,b,e,c;q):=&\sum_{h\neq 0}G_{\overline{\chi}}(h;c)\sum_{\pm}\sum_{w=1}^{\infty}\frac{(ae,w)r_w(h)}{w^2}\\ &\times\int_0^\infty L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x,\pm(h-aex);c\right)\,dx.\end{aligned} \tag{4.15}$$

The integrands in the two terms $\tilde{\mathcal{S}}^M$ and $\mathcal{S}^M$ differ only for $x=h/(ae)+O(1/(ae))$. Since by (4.9) (cf. (6.2) below) the h-sum in both terms is essentially over $1\leqslant|h|\leqslant eq^{1+\varepsilon}/(ab)$, the contribution of their difference to (4.11) is at most (cf. (2.25), (2.30), (4.9), (4.10))

$$\ll_\varepsilon q^\varepsilon\sum_{ae|\ell}\sum_{q|c}\frac{1}{c^2}\sum_{1\leqslant h\leqslant eq^{1+\varepsilon}}c^{1/2}(h,c)^{1/2}\sum_{w=1}^{\infty}\frac{(ae,w)(h,w)}{w^2}\left(\frac{q}{aeh}\right)^{1/2}\ll_\varepsilon q^{3\varepsilon-1/2}. \tag{4.16}$$

5. The main term

In this section, we will evaluate the contribution of the term (4.15) to (4.11):

$$\begin{aligned}&\sum_{abe=\ell}\frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}}\sum_{w\geqslant 1}\frac{(ae,w)}{w^2}\\ &\times\sum_{q|c}\frac{1}{c^2}\sum_{h\neq 0}r_w(h)G_{\overline{\chi}}(h;c)\sum_{\pm}\int_0^\infty L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x,\pm(h-aex);c\right)\,dx.\end{aligned} \tag{5.1}$$

More precisely, we shall first evaluate the c- and h-sums above then average trivially over a, b, e, w.

To do so we proceed essentially as in [KMV, pp. 117–122]. We substitute the definition (4.7) of $g^{\pm}$ and make a change of variables

$$\xi := \frac{|h|}{c^2} y, \qquad \eta := \frac{ae}{|h|} x$$

in order to remove all parameters from the oscillating functions. Secondly, we replace the negative values of h in (5.1) (which only contribute to the "$-$" case in $\sum_{\pm}$) by their absolute values. To simplify the notation, let us write (cf. (2.25))

$$\mathcal{L}(\eta) := L_w(h\eta) = \log \eta + 2\gamma + \log\left(\frac{h}{w^2}\right) =: \log \eta + \Lambda,$$

say. Then the c, h-sum in (5.1) equals

$$\begin{aligned}(5.2)\quad & \frac{1}{\sqrt{ae}} \sum_{q|c} \frac{1}{c} \sum_{h \geqslant 1} r_w(h) G_{\overline{\chi}}(h; c) \int_0^\infty \int_0^\infty \phi(4\pi\sqrt{\xi\eta}) \\ & \times \Big\{ \delta_{\eta<1} \mathcal{L}(1-\eta) \mathcal{J}^+(\sqrt{(1-\eta)\xi}) + \delta_{\eta>1} \mathcal{L}(\eta-1) \mathcal{J}^-(\sqrt{(\eta-1)\xi}) + \chi(-1) \mathcal{L}(\eta+1) \mathcal{J}^-(\sqrt{(\eta+1)\xi}) \Big\} \\ & \times K_{(ae,w),w}\left(\frac{h\eta}{ae}\right) W_1\left(\frac{c^2\xi}{hq}\right) W_2\left(\frac{abh\eta}{eq}\right) \frac{d\xi d\eta}{(\xi\eta)^{1/2}}.\end{aligned}$$

Let us also write

$$X_w(\eta) := K_{(ae,w),w}\left(\frac{q\eta}{a^2 b}\right) W_2(\eta).$$

Its Mellin transform $\widehat{X}_w$ satisfies essentially the same properties as $\widehat{W}_2$. To see this, observe first that by (4.4), W_2 is up to a negligible error supported on $[0, q^\varepsilon]$, so we can replace $K_{(ae,w),w}\left(q\eta/(a^2 b)\right)$ by

$$K_w^*(\eta) := K_{(ae,w),w}\left(\frac{q\eta}{a^2 b}\right) \omega\left(\frac{\eta}{q^\varepsilon}\right),$$

where, as usual, ω is a smooth cut-off function. Then, by (2.30), (4.12), and sufficiently many integrations by parts, we find that

$$\widehat{K}_w^*(u) = \int_0^\infty K_w^*(\eta) \eta^{u-1} \, d\eta \ll_{j, \mathfrak{Re}\, u} q^\varepsilon |u|^{-j}$$

for $\mathfrak{Re}\, u > 0$ and any $j \geqslant 0$. Finally, by (3.14),

$$\widehat{X}_w(u) = \frac{1}{2\pi i} \int_{(\frac{1}{2}\mathfrak{Re}\, u)} \widehat{K}_w^*(u-v) \widehat{W}_2(v) \, dv \ll_{j,\varepsilon} q^\varepsilon |u|^{-j}$$

for $\varepsilon \leqslant \mathfrak{Re}\, u \leqslant 5$, say.

Our next aim is to transform the double integral in (5.2) by several applications of Mellin's inversion formula: using (2.26) and (2.17), we write $\mathcal{J}^{\pm}$ and ϕ as inverse Mellin transforms. Then the ξ, η-integral in (5.2) equals

$$\begin{aligned}(5.3)\quad & \int_0^\infty \int_0^\infty \frac{4}{(2\pi i)^2} \int_{(0.2)} \int_{(*)} \phi^*(u_1) (2\pi\sqrt{\xi\eta})^{1+2u_1} (2\pi)^{-2u_2} \Gamma(u_2)^2 \xi^{-u_2} \\ & \times \left(\frac{\delta_{\eta<1} \mathcal{L}(1-\eta)}{(1-\eta)^{u_2}} + \frac{\delta_{\eta>1} \mathcal{L}(\eta-1) \cos \pi u_2}{(\eta-1)^{u_2}} + \frac{\chi(-1) \mathcal{L}(\eta+1) \cos \pi u_2}{(\eta+1)^{u_2}} \right) du_2 \, du_1 \\ & \times W_1\left(\frac{c^2\xi}{hq}\right) X_w\left(\frac{abh\eta}{eq}\right) \frac{d\xi d\eta}{(\xi\eta)^{1/2}}.\end{aligned}$$

Since the u_1-, u_2- and ξ-integrals are absolutely convergent (using (4.4)), we can pull the ξ-integration inside and calculate it explicitly in terms of the Mellin transform $\widehat{W}_1$ of W_1. Then we write X_w as

an inverse Mellin transform getting that (5.3) equals

$$\int_0^\infty \frac{4}{(2\pi i)^2}\int_{(0.2)}\int_{(0.6)} \phi^*(u_1)(2\pi)^{1+2u_1-2u_2}\eta^{1/2+u_1}\Gamma(u_2)^2$$
$$\times\left(\frac{\delta_{\eta<1}\mathcal{L}(1-\eta)}{(1-\eta)^{u_2}}+\frac{\delta_{\eta>1}\mathcal{L}(\eta-1)\cos\pi u_2}{(\eta-1)^{u_2}}+\frac{\chi(-1)\mathcal{L}(\eta+1)\cos\pi u_2}{(\eta+1)^{u_2}}\right)\left(\frac{c^2}{hq}\right)^{-1-u_1+u_2}$$
$$\times\widehat{W}_1\left(1+u_1-u_2\right)\,du_2\,du_1\,\frac{1}{2\pi i}\int_{(0.9)}\left(\frac{abh\eta}{eq}\right)^{-u_3}\widehat{X}_w(u_3)\,du_3\,\frac{d\eta}{\eta^{1/2}}.$$

Here we shifted the u_2-integration to $\mathfrak{Re}\,u_2 = 0.6$ since $\widehat{W_1}(u)$ is rapidly decaying on the line $\mathfrak{Re}\,u = 0.6$. Since again all integrals are absolutely convergent, we can pull the η-integration inside and calculate the three terms explicitly (as in [KMV, (38)]) using [GR, 3.191.1, 3.191.2, 3.194.3]. We find

$$\int_0^\infty\left(\frac{\delta_{\eta<1}\mathcal{L}(1-\eta)}{(1-\eta)^{u_2}}+\frac{\delta_{\eta>1}\mathcal{L}(\eta-1)\cos\pi u_2}{(\eta-1)^{u_2}}+\frac{\chi(-1)\mathcal{L}(\eta+1)\cos\pi u_2}{(\eta+1)^{u_2}}\right)\eta^{1+u_1-u_3}\frac{d\eta}{\eta}$$
$$=\cos(\pi u_2)(-\partial_{u_2}+\Lambda)\frac{\Gamma(1+u_1-u_3)\Gamma(-1-u_1+u_3+u_2)}{\Gamma(u_2)}\left(\chi(-1)+\frac{\sin(\pi(u_3-u_1))}{\sin(\pi u_2)}\right)$$
$$-(-\partial_{u_2}+\Lambda)\frac{\Gamma(1+u_1-u_3)\Gamma(-1-u_1+u_3+u_2)}{\Gamma(u_2)}\frac{\sin(\pi(-u_1+u_3+u_2))}{\sin(\pi u_2)}.$$

Introducing the new variable $u_4 := 1+u_1-u_2$ as a substitute for u_2, we see that (5.3) equals

$$\begin{aligned}&\frac{-4}{(2\pi i)^3}\int_{(0.2)}\int_{(0.6)}\int_{(0.9)}\phi^*(u_1)(2\pi)^{2u_4-1}\Gamma(1+u_1-u_4)^2\Gamma(1+u_1-u_3)\\&\times\left(\frac{c^2}{hq}\right)^{-u_4}\widehat{W}_1\left(u_4\right)\left(\frac{abh}{eq}\right)^{-u_3}\widehat{X}_w(u_3)\\&\times\left\{\cos(\pi(u_1-u_4))(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\left(\chi(-1)+\frac{\sin(\pi(u_3-u_1))}{\sin(\pi(u_4-u_1))}\right)\right.\\&\qquad\left.-(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\frac{\sin(\pi(u_3-u_4))}{\sin(\pi(u_4-u_1))}\right\}\,du_3\,du_4\,du_1.\end{aligned}\tag{5.4}$$

Using the identities

$$(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)} = \frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4)-\frac{\Gamma'}{\Gamma}(u_3-u_4)+\Lambda\right),$$

$$\frac{\sin(\pi(u_3-u_4))}{\sin(\pi(u_4-u_1))} = -\cos(\pi(u_1-u_3))+\cos(\pi(u_1-u_4))\frac{\sin(\pi(u_3-u_1))}{\sin(\pi(u_4-u_1))},$$

it is straightforward to verify that the last two lines in (5.4) can be simplified to

$$\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\Big\{-\pi\sin(\pi(u_1-u_3))+$$
$$\Big(\cos(\pi(u_1-u_3))+\chi(-1)\cos(\pi(u_1-u_4))\Big)\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4)-\frac{\Gamma'}{\Gamma}(u_3-u_4)+\Lambda\right)\Big\}.$$

In particular, we observe that the triple integral is absolutely convergent (since ϕ^*, $\widehat{W_1}$, and $\widehat{X}_w$ are sufficiently nice) and the integrand is holomorphic whenever $0 < \mathfrak{Re}\,u_4 < \mathfrak{Re}\,u_3 < 1+\mathfrak{Re}\,u_1$. Let us shift the u_4-contour to $\mathfrak{Re}\,u_4 = \varepsilon\,(<0.1)$ and the u_3-contour to $\mathfrak{Re}\,u_3 = 1.1$.

We now substitute this triple integral back into (5.2) and perform the (absolutely convergent) sum over c and h. To justify this, we need to evaluate for $s = 1+2u_4$, $t = u_3-u_4$ the Dirichlet

series

$$D_{w,q}(\chi,s,t):=\sum_{q|c}\frac{1}{c^s}\sum_{h\geqslant 1}\frac{G_{\overline{\chi}}(h;c)r_w(h)}{h^t}.$$

First we need to compute the Gauß sum $G_{\overline{\chi}}(h;c)$: we denote by q^* the conductor of χ and, slightly abusing notation, we write χ also for the primitive character of modulus q^* underlying χ mod q. For $q\mid c$ we consider the unique factorization $c=q^*q_1q_2c_1c_2$ where $q=q^*q_1q_2$, $c_1q_1\mid (q^*)^\infty$ and $(c_2q_2,q^*)=1$. Then

$$G_{\overline{\chi}}(h;c)=\overline{\chi}(c_2q_2)G_{\overline{\chi}}(h;q^*q_1c_1)r_{c_2q_2}(h)$$

(with $r_{c_2q_2}(h)$ being the Ramanujan sum, cf. (2.29)). Moreover, $G_{\overline{\chi}}(h;q^*q_1c_1)=0$ unless $c_1q_1\mid h$ in which case

$$G_{\overline{\chi}}(h;q^*q_1c_1)=\chi\left(\frac{h}{c_1q_1}\right)c_1q_1G_{\overline{\chi}}(1;q^*).$$

Summarizing the above computation, one has

$$G_{\overline{\chi}}(h;c)=\delta_{c_1q_1|h}\overline{\chi}(c_2q_2)\chi\left(\frac{h}{c_1q_1}\right)c_1q_1r_{c_2q_2}(h)G_{\overline{\chi}}(1;q^*). \tag{5.5}$$

Therefore

$$D_{w,q}(\chi,s,t)=\frac{\overline{\chi}(q_2)G_{\overline{\chi}}(1;q^*)}{(q^*)^sq_1^{s+t-1}q_2^s}\sum_{c_1|(q^*)^\infty}\frac{1}{c_1^{s+t-1}}\sum_{(c_2,q^*)=1}\frac{\overline{\chi}(c_2)}{c_2^s}\sum_{h\geqslant 1}\frac{\chi(h)r_w(c_1q_1h)r_{c_2q_2}(h)}{h^t}.$$

For $\sigma:=\mathfrak{Re}\,s$, and $\tau:=\mathfrak{Re}\,t$ sufficiently large the c_1,c_2,h-sum factors as an Euler product over the primes:

$$\sum_{c_1|(q^*)^\infty}\frac{1}{c_1^{s+t-1}}\sum_{(c_2,q^*)=1}\frac{\overline{\chi}(c_2)}{c_2^s}\sum_{h\geqslant 1}\frac{\chi(h)r_w(c_1q_1h)r_{c_2q_2}(h)}{h^t}=\prod_p\Pi_p(\chi,s,t), \tag{5.6}$$

say. We collected some useful properties of the Euler factors $\Pi_p(\chi,s,t)$ in Lemma 2 at the end of this section. These properties imply that for $\mathfrak{Re}\,u_4=\varepsilon\,(<0.1)$ and $\mathfrak{Re}\,u_3=1.1$ the series $D_{w,q}(\chi,s,t)$ is absolutely convergent and in the domain $\sigma>1$, $\tau>0$ it decomposes as

$$D_{w,q}(\chi,s,t)=\zeta(s+t-1)L(\chi,t)H_{w,q}(\chi,s,t), \tag{5.7}$$

where $H_{w,q}(\chi,s,t)$ is a holomorphic function. Moreover, for $0<\varepsilon<0.1$,

$$\mathfrak{Re}\,s=1+2\varepsilon,\qquad \varepsilon/2<\mathfrak{Re}\,t<3\varepsilon/2,$$

one has

$$H_{w,q}(\chi,s,t)\ll_\varepsilon q^\varepsilon(q_1,w)w^{1-\varepsilon/3}(q^*)^{-1/2}. \tag{5.8}$$

Using

$$\Lambda=2\gamma-\log\bigl(w^2\bigr)+\log(h)$$

we obtain that (5.2) equals

$$\frac{1}{\sqrt{ae}}\frac{-4}{(2\pi i)^3}\int_{(0.2)}\int_{(\varepsilon)}\int_{(1.1)}\phi^*(u_1)(2\pi)^{2u_4-1}\Gamma(1+u_1-u_4)\Gamma(1+u_1-u_3)\Gamma(u_3-u_4)$$
$$\times q^{u_4}\widehat{W}_1(u_4)\left(\frac{ab}{eq}\right)^{-u_3}\widehat{X}_w(u_3)\left\{\sum_{j=0}^1\partial_{u_3}^j\bigl(\zeta(u_3+u_4)\widetilde{L}(u_3,u_4)\bigr)F_j\right\}du_3\,du_4\,du_1,$$

where

$$\widetilde{L}(u_3,u_4):=L(\chi,u_3-u_4)H_{w,q}(\chi,1+2u_4,u_3-u_4)$$

and

$$F_0(u_1,u_3,u_4) := -\pi\sin(\pi(u_1-u_3)) + \Big(\cos(\pi(u_1-u_3)) + \chi(-1)\cos(\pi(u_1-u_4))\Big) \times \left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_3-u_4) + 2\gamma - \log\big(w^2\big)\right),$$

$$F_1(u_1,u_3,u_4) := -\Big(\cos(\pi(u_1-u_3)) + \chi(-1)\cos(\pi(u_1-u_4))\Big).$$

Let us now shift the u_3-contour from $\mathfrak{Re}\, u_3 = 1.1$ to $\mathfrak{Re}\, u_3 = 2\varepsilon$; we will show below that there is no pole at $u_3+u_4=1$. Then $\mathfrak{Re}\,(u_3-u_4)=\varepsilon$, hence

$$\partial_{u_3}^j L(\chi, u_3-u_4) \ll_{j,\varepsilon} (q^*)^{1/2}$$

by the functional equation for $L(\chi,t)$ with implied constants depending on j, ε and (polynomially) on $|\mathfrak{Im}\,(u_3-u_4)|$. In addition, (5.8) combined with Cauchy's integral formula shows that

$$\partial_{u_3}^j H_{w,q}(\chi, 1+2u_4, u_3-u_4) \ll_{j,\varepsilon} q^{\varepsilon}(q_1,w) w^{1-\varepsilon/3}(q^*)^{-1/2},$$

therefore (5.2) summed over $abe=\ell$ is bounded by

$$\ll_\varepsilon q^{10\varepsilon}(q_1,w)\log(w) w^{1-\varepsilon/3} \sum_{abe=\ell} \frac{(e/ab)^{\mathfrak{Re}\, u_3}}{a\sqrt{abe}} \ll_\varepsilon q^{12\varepsilon}(q_1,w) w^{1-\varepsilon/4}\ell^{-1/2}.$$

Finally, averaging over w the above bound against the weight $(ae,w)/w^2$, we obtain that the main term (5.1) is bounded by

$$\ll_\varepsilon q^{13\varepsilon}\ell^{-1/2}. \tag{5.9}$$

To conclude the analysis of the main term, it remains to show that the pole of the zeta-function at $u_3+u_4=1$ does not contribute anything. Let us only focus on the factors depending on u_3:

$$G(u_1,u_3,u_4) := \Gamma(1+u_1-u_3)\Gamma(u_3-u_4)\left(\frac{ab}{eq}\right)^{-u_3} \widehat{X}_w(u_3) \left\{\sum_{j=0}^{1} \partial_{u_3}^j\big(\zeta(u_3+u_4)\widetilde{L}(u_3,u_4)\big) F_j\right\}.$$

If R_j denotes the contribution of the j-term to the residue of $G(u_1,u_3,u_4)$ at $u_3=1-u_4$, then

$$R_0 = \Gamma(u_1+u_4)\Gamma(1-2u_4)\left(\frac{ab}{eq}\right)^{u_4-1} \widehat{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4) \times \left\{ \begin{array}{l} +\pi\sin(\pi(u_1+u_4)) + \\ \left\{ \begin{array}{l} +2\sin(\pi u_1)\sin(\pi u_4) \\ -2\cos(\pi u_1)\cos(\pi u_4) \end{array} \right\} \left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(1-2u_4) + 2\gamma - \log\big(w^2\big)\right) \end{array} \right\},$$

$$R_1 = \Gamma(u_1+u_4)\Gamma(1-2u_4)\left(\frac{ab}{eq}\right)^{u_4-1} \widehat{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4) \times \left\{ \begin{array}{l} -\pi\sin(\pi(u_1+u_4)) + \\ \left\{ \begin{array}{l} +2\sin(\pi u_1)\sin(\pi u_4) \\ -2\cos(\pi u_1)\cos(\pi u_4) \end{array} \right\} \left(-\frac{\Gamma'}{\Gamma}(u_1+u_4) + \frac{\Gamma'}{\Gamma}(1-2u_4) + \frac{\widehat{X}_w'}{\widehat{X}_w}(1-u_4) - \log\left(\frac{ab}{eq}\right)\right) \end{array} \right\}.$$

Here the upper line corresponds to $\kappa=0$ and the lower line to $\kappa=1$, and we have used $\chi(-1)=(-1)^\kappa$. Altogether the residual integral equals, after shifting the u_1-integration to $(-\varepsilon/2)$ and interchanging the u_1- and u_4-integration,

$$\begin{aligned} &\frac{8}{(2\pi i)^2}\int_{(\varepsilon)}\int_{(-\varepsilon/2)} \phi^*(u_1)(2\pi)^{2u_4-1}\Gamma(1+u_1-u_4)\Gamma(u_1+u_4)\Gamma(1-2u_4) \\ &\times q^{u_4}\widehat{W}_1(u_4)\left(\frac{ab}{eq}\right)^{u_4-1} \widehat{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4) \\ &\times \left\{ \begin{array}{l} -\sin(\pi u_1)\sin(\pi u_4) \\ +\cos(\pi u_1)\cos(\pi u_4) \end{array} \right\} \left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_1+u_4) + \widetilde{\Lambda}(u_4)\right) du_1\, du_4, \end{aligned} \tag{5.10}$$

where

$$\widetilde{\Lambda}(u_4) := \frac{\widehat{X}_w'}{\widehat{X}_w}(1-u_4) + 2\gamma - \log\left(\frac{abw^2}{eq}\right).$$

We recast the inner integral as

$$\frac{1}{2\pi i}\int_{(-\varepsilon/2)} \phi^*(u_1)\left(\widetilde{\Lambda}(u_4) - \partial_{u_4}\right)\Gamma(1+u_1-u_4)\Gamma(u_1+u_4)\left\{\begin{array}{l} -\sin(\pi u_1) \\ +\cos(\pi u_1)\end{array}\right\} du_1$$

and use (2.18) to see that (5.10) equals

$$\frac{8\pi q}{2\pi i}\int_{(\frac12)} (2\pi)^{2u_4-1}\Gamma(1-2u_4)\widehat{W}_1(u_4)\left(\frac{ab}{e}\right)^{u_4-1}\widehat{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4)$$
$$\times\left\{\begin{array}{l}\sin(\pi u_4)\\ \cos(\pi u_4)\end{array}\right\}\left(\widetilde{\Lambda}(u_4)-\partial_{u_4}\right)\tilde{\phi}\left(i\left(\tfrac12-u_4\right)\right)\left\{\begin{array}{l}1\\ \frac{\cot(\pi u_4)}{i(\frac12-u_4)}\end{array}\right\} du_4.$$

If $\phi = J_{k-1}$, $k \equiv \kappa \pmod 2$ then the integral vanishes by $\tilde{\phi} = 0$. Otherwise we shift ∂_{u_4} to the other factors by partial integration. Then we sum over ν as in (4.1) and recall that, by the definition of ϕ and W_1,

$$\widehat{W}_1(u_4) = u_4^{2\nu}\widehat{\overline{W}}(u_4) \qquad \text{and} \qquad \tilde{\phi}(t) = \tilde{\phi}_0(t)\overline{\alpha_\nu(t)}\alpha_\xi(t).$$

For $t \in \mathbb{R}$ we have $\alpha_\nu(t) \in \mathbb{R}$, hence the sum over ν introduces factors

$$\sum_{\nu=0}^4 \alpha_\nu\left(i\left(\tfrac12-u_4\right)\right)u_4^{2\nu} \qquad \text{or} \qquad \sum_{\nu=0}^4 \alpha_\nu\left(i\left(\tfrac12-u_4\right)\right)\frac{\partial}{\partial u_4}u_4^{2\nu}$$

to each term. By (3.8)–(3.9) these factors vanish, that is, the residual integral (5.10) is zero in all cases. This completes the analysis of the main term.

Without the additional zeros in the approximate functional equation, we might still succeed at the cost of much more work. Applying the functional equation of $L(s,\chi)$, expressing $K_{(ae,w),w}(y)$ in terms of $L_{\frac{w}{(ae,w)}}(y)$ and therefore X_w in terms of W, it should be possible to see that the polar contribution (5.10) resembles exactly the contribution of the cusps $\mathfrak{a} = 0, \infty$ of $\mathcal{Q}(\ell)$, see (3.19).

We conclude this section by stating and proving some useful properties for the Euler factors $\Pi_p(\chi,s,t)$ in (5.6).

Lemma 2. *Let $\sigma = \mathfrak{Re}\, s > 1$ and $\tau = \mathfrak{Re}\, t > 0$. For a prime p let v_p denote the p-adic valuation, and let ζ_p (resp. L_p) denote the corresponding Euler factor of the Riemann zeta function (resp. Dirichlet L-function).*

a) For $(p, qw) = 1$,

$$\Pi_p(\chi,s,t) = \zeta_p(s+t-1)\frac{L_p(\chi,t)}{L_p(\overline{\chi},s)}.$$

b) For $p \mid q^*$,

$$|\Pi_p(\chi,s,t)| \leqslant 3p^{\min(v_p(q_1),v_p(w))+(1-\tau)v_p(w)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

c) For $(p, q^*) = 1$, $p \mid qw$,

$$|\Pi_p(\chi,s,t)| \leqslant 4p^{v_p(q_2)+(1-\tau)v_p(w)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

Proof. a) For $(p, qw) = 1$ we use the notation

$$\alpha := v_p(c_2), \qquad \beta := v_p(h)$$

in the sum (5.6), then

$$\begin{aligned}\Pi_p(\chi,s,t) &= \sum_{\alpha=0}^{\infty}\frac{\overline{\chi}(p^\alpha)}{p^{\alpha s}}\sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)r_{p^\alpha}(p^\beta)}{p^{\beta t}}\\ &= \sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)}{p^{\beta t}}\left(1+\sum_{\alpha=1}^{\beta}\frac{\overline{\chi}(p^\alpha)}{p^{\alpha s}}(p^\alpha-p^{\alpha-1})-\frac{\overline{\chi}(p^{\beta+1})}{p^{(\beta+1)s}}p^\beta\right)\\ &= \frac{1-\overline{\chi}(p)p^{-s}}{1-\overline{\chi}(p)p^{1-s}}\sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)}{p^{\beta t}}\left(1-\frac{\overline{\chi}(p^{\beta+1})}{p^{(\beta+1)s}}p^{\beta+1}\right)\\ &= \frac{1-\overline{\chi}(p)p^{-s}}{1-\overline{\chi}(p)p^{1-s}}\left(\frac{1}{1-\chi(p)p^{-t}}-\frac{\overline{\chi}(p)p^{1-s}}{1-p^{1-s-t}}\right)\\ &= \frac{1-\overline{\chi}(p)p^{-s}}{(1-\chi(p)p^{-t})(1-p^{1-s-t})}.\end{aligned}$$

b) For $p \mid q^*$ we use the notation

$$\alpha := v_p(c_1), \qquad \beta := v_p(h), \qquad \gamma := v_p(q_1), \qquad \delta := v_p(w)$$

in the sum (5.6), then clearly

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{\left|r_{p^\delta}(p^{\alpha+\beta+\gamma})\right|}{p^{\beta\tau}}.$$

We distinguish between two cases. For $\gamma \geqslant \delta$ we infer

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{p^\delta}{p^{\beta\tau}} = p^\delta\zeta_p(\sigma+\tau-1)\zeta_p(\tau).$$

For $\gamma < \delta$ we infer

$$\begin{aligned}|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\delta-\gamma-1}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\left(\frac{p^{\delta-1}}{p^{(\delta-\gamma-1-\alpha)\tau}}+\sum_{\beta=\delta-\gamma-\alpha}^{\infty}\frac{p^\delta}{p^{\beta\tau}}\right)+\sum_{\alpha=\delta-\gamma}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{p^\delta}{p^{\beta\tau}}\\ &= p^\gamma\sum_{\alpha=0}^{\delta-\gamma-1}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\left(\frac{p^{\delta-\gamma-1}}{p^{(\delta-\gamma-1-\alpha)\tau}}+\frac{p^{\delta-\gamma}}{p^{(\delta-\gamma-\alpha)\tau}}\zeta_p(\tau)\right)+p^\delta\zeta_p(\tau)\sum_{\alpha=\delta-\gamma}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\\ &\leqslant 2p^{\gamma+(\delta-\gamma)(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma-1)+p^{\gamma+(\delta-\gamma)(2-\sigma-\tau)}\zeta_p(\tau)\zeta_p(\sigma+\tau-1).\end{aligned}$$

In both cases we conclude

$$|\Pi_p(\chi,s,t)| \leqslant 3p^{\min(\gamma,\delta)+\delta(1-\tau)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

c) For $(p,q^*)=1$, $p \mid qw$, we use the notation

$$\alpha := v_p(c_2), \qquad \beta := v_p(h), \qquad \gamma := v_p(q_2), \qquad \delta := v_p(w)$$

in the sum (5.6), then clearly

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha\sigma}}\sum_{\beta=0}^{\infty}\frac{\left|r_{p^\delta}(p^\beta)r_{p^{\alpha+\gamma}}(p^\beta)\right|}{p^{\beta\tau}}.$$

We distinguish between two cases. For $\gamma \geqslant \delta$ we infer (note that $\gamma > 0$ in this case)

$$\begin{aligned}|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\infty} \frac{p^\delta}{p^{\alpha\sigma}} \left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}} + \sum_{\beta=\alpha+\gamma}^{\infty} \frac{p^{\alpha+\gamma}}{p^{\beta\tau}} \right) \\ &\leqslant p^\delta \zeta_p(\tau) \sum_{\alpha=0}^{\infty} \frac{1}{p^{\alpha\sigma}} \left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}} + \frac{p^{\alpha+\gamma}}{p^{(\alpha+\gamma)\tau}} \right) \\ &\leqslant 2p^{\delta+\gamma(1-\tau)} \zeta_p(\tau)\zeta_p(\sigma+\tau-1) \\ &\leqslant 2p^{\gamma+\delta(1-\tau)} \zeta_p(\tau)\zeta_p(\sigma+\tau-1).\end{aligned}$$

For $\gamma < \delta$ we infer

$$\begin{aligned}|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\delta-\gamma-1} \frac{p^{\alpha+\gamma}}{p^{\alpha\sigma}} \left(\frac{p^{\delta-1}}{p^{(\delta-1)\tau}} + \sum_{\beta=\delta}^{\infty} \frac{p^\delta}{p^{\beta\tau}} \right) + \sum_{\alpha=\delta-\gamma}^{\infty} \frac{p^\delta}{p^{\alpha\sigma}} \left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}} + \sum_{\beta=\alpha+\gamma}^{\infty} \frac{p^{\alpha+\gamma}}{p^{\beta\tau}} \right) \\ &\leqslant 2p^{\gamma+\delta(1-\tau)} \zeta_p(\tau)\zeta_p(\sigma-1) + 2p^{\delta+\gamma(1-\tau)} \zeta_p(\tau) \sum_{\alpha=\delta-\gamma}^{\infty} \frac{1}{p^{\alpha(\sigma+\tau-1)}} \\ &= 2p^{\gamma+\delta(1-\tau)} \zeta_p(\tau)\zeta_p(\sigma-1) + 2p^{\delta-\sigma(\delta-\gamma)+\delta(1-\tau)} \zeta_p(\tau)\zeta_p(\sigma+\tau-1).\end{aligned}$$

In both cases we conclude

$$|\Pi_p(\chi,s,t)| \leqslant 4p^{\gamma+\delta(1-\tau)} \zeta_p(\sigma-1)\zeta_p(\tau).$$

The proof of Lemma 2 is complete □

6. Transforming the Kloosterman sums

6.1. **Applying the trace formula.** Finally we estimate the contribution of (4.14) to (4.11). This time, we fix c and evaluate the h-sum non-trivially: in other words, we will bound the terms (4.14) $\mathcal{S}^{E,\pm}(a,b,e,c;q)$ for c satisfying (cf. (4.9))

$$q \mid c, \qquad q \leqslant c \leqslant \frac{\sqrt{e}q^{1+\varepsilon}}{\sqrt{ab}}.$$

As a first step, we use the identity

$$\sum_{w \geqslant 1} F(w,(ae,w)) = \sum_{r \mid ae} \sum_{(ae,w)=r} F(w,r) = \sum_{rs \mid ae} \mu(s) \sum_{w \equiv 0 \,(rs)} F(w,r)$$

and write (4.14) as

$$\begin{aligned}&\sum_{\pm} \sum_{rs \mid ae} r\mu(s) \sum_{w \equiv 0 \,(rs)} \frac{1}{w^2} \sum_{h \neq 0} G_{\overline{\chi}}(h;c) \sum_{n=1}^{\infty} \tau(n) S\left(\mp h, \pm n; w\right) \\ &\times \int_0^\infty \mathcal{J}^\pm \left(\frac{\sqrt{n(\pm(h-aex))}}{w} \right) K_{r,w}(x) g^\pm(x, \pm(h-aex); c) \psi(\pm(h-aex))\, dx.\end{aligned}$$

We want to apply the trace formulae (2.11) and (2.12) to the w-sum. This needs some preparation. By (4.9) we can restrict the x-integration to

$$|h - aex| \leqslant aexq^\varepsilon \leqslant \frac{eq^{1+\varepsilon}}{ab} \tag{6.1}$$

and the h-summation to

$$|h| \leqslant \frac{eq^{1+\varepsilon}}{ab}, \tag{6.2}$$

up to negligible error. Let ρ be a smooth nonnegative function with bounded derivatives, supported on $[1/2,2]$ such that $\rho(y)+\rho(2y)=1$ for $y\in[1/2,1]$. Then $\sum_{\nu\in\mathbb{Z}}\rho(2^\nu y)=1$ for $y>0$. We apply this smooth partition of unity to all variables and insert (4.7); thus we will bound $O(\log^6 q)$ terms ((6.4), (6.6), (6.9) show that each of W, H, N, R, X, Y can be taken from the interval $[1/2,\ell^3q^{1+\varepsilon}]$), of the shape

$$\begin{aligned}&\sum_{rs|ae}r\mu(s)\sum_{w\equiv 0\,(rs)}\frac{\rho(w/W)}{w^2}\sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sum_n\rho\left(\frac{n}{N}\right)\tau(n)S(\mp h,\pm n;w)\\ &\times\int_0^\infty\int_0^\infty K_{r,w}(x)\rho\left(\frac{\pm(h-aex)}{R}\right)\rho\left(\frac{x}{X}\right)\rho\left(\frac{y}{Y}\right)W_1\left(\frac{y}{q}\right)W_2\left(\frac{a^2bx}{q}\right)\\ &\times\mathcal{J}^{\pm}\left(\frac{\sqrt{n(\pm(h-aex))}}{w}\right)\mathcal{J}^{\pm}\left(\frac{\sqrt{\pm(h-aex)y}}{c}\right)\phi\left(\frac{4\pi\sqrt{aexy}}{c}\right)\frac{dydx}{(xy)^{1/2}}.\end{aligned}\tag{6.3}$$

(More precisely, for $1/2\leqslant R\leqslant 1$ we adjust the first ρ-factor by the function ψ.) In view of (6.1), (4.9), (4.4), (2.31) and (4.12), and the remark following (4.8), we can assume

$$\frac{1}{2}\leqslant R\leqslant aeXq^{\varepsilon},\qquad \frac{q^{1-\varepsilon}}{ae}\leqslant X\leqslant\frac{q^{1+\varepsilon}}{a^2b},\qquad \frac{abq^{1-\varepsilon}}{e}\leqslant Y\leqslant q^{1+\varepsilon},\qquad \frac{1}{2}\leqslant W\leqslant\frac{rq^{1/2+\varepsilon}}{a\sqrt{b}}.\tag{6.4}$$

Now we use (2.16) to integrate the first factor in the third line of (6.3) by parts sufficiently many times; in order to apply (2.16) we change variables $\mathfrak{r}:=\pm(h-aex)\asymp R$. By (2.30) and (4.4), the j-th derivative with respect to $\mathfrak{r}$ of the integrand without the $\mathcal{J}^{\pm}(\sqrt{n\mathfrak{r}}/w)$ factor is

$$\ll_{\varepsilon,j}q^{\varepsilon}\left(\frac{1}{R}+\frac{1}{Xae}+\frac{\sqrt{Y}}{c\sqrt{R}}+\frac{\sqrt{Y}}{c\sqrt{Xae}}\right)^j\ll_{\varepsilon,j}q^{\varepsilon}\left(\frac{1}{R}+\frac{\sqrt{Y}}{q\sqrt{R}}\right)^j.$$

This shows, by (2.16), that the integral in (6.3) is negligible unless

$$\frac{W}{\sqrt{N}}\left(\frac{1}{\sqrt{R}}+\frac{\sqrt{Y}}{q}\right)\geqslant q^{-\varepsilon}.\tag{6.5}$$

Note that this implies either $\sqrt{RN}/W\leqslant q^{\varepsilon}$ or $\sqrt{N}/W\leqslant q^{-1/2+\varepsilon}$ (since $Y\leqslant q^{1+\varepsilon}$), and so in any case

$$\frac{\sqrt{RN}}{W}\leqslant\sqrt{e}q^{\varepsilon}.\tag{6.6}$$

Let us now define

$$\begin{aligned}\Psi(h,n;z):=&\frac{z\rho(n/N)}{4\pi\sqrt{|h|n}}\rho\left(\frac{4\pi\sqrt{|h|n}}{zW}\right)\int_0^\infty\int_0^\infty K_{r,4\pi\sqrt{|h|n}/z}(x)\\ &\times\rho\left(\frac{\pm(h-aex)}{R}\right)\rho\left(\frac{x}{X}\right)\rho\left(\frac{y}{Y}\right)W_1\left(\frac{y}{q}\right)W_2\left(\frac{a^2bx}{q}\right)\\ &\times\mathcal{J}^{\pm}\left(z\frac{\sqrt{\pm(h-aex)}}{4\pi\sqrt{|h|}}\right)\mathcal{J}^{\pm}\left(\frac{\sqrt{\pm(h-aex)y}}{c}\right)\phi\left(\frac{4\pi\sqrt{aexy}}{c}\right)\frac{dydx}{(xy)^{1/2}}.\end{aligned}\tag{6.7}$$

Then (6.3) equals

$$\sum_{rs|ae}r\mu(s)\sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sum_n\tau(n)\sum_{w\equiv 0\,(rs)}\frac{1}{w}S(\mp h,\pm n;w)\Psi\left(h,n;\frac{4\pi\sqrt{|h|n}}{w}\right).\tag{6.8}$$

We are now in a position to apply Kuznetsov's trace formula (2.11)–(2.12) for

$$\text{level } rs,\ \text{trivial nebentypus and weight } 0.$$

The innermost sum in (6.8) equals

$$\sum_{f\in\mathcal{B}(rs,\chi_0)} \tilde\Psi(h,n;t_f)\frac{4\pi\sqrt{|h|n}}{\cosh(\pi t_f)}\overline{\rho_f}(|h|)\rho_f(n) + \text{ two similar terms}$$

corresponding to holomorphic forms and Eisenstein series (or a similar expression with $\check\Psi$ in place of $\tilde\Psi$). We substitute this into (6.8), and are left with bounding

$$\sum_{rs|ae} r\mu(s) \sum_{f\in\mathcal{B}(rs,\chi_0)} \sum_h G_{\overline\chi}(h;c)\rho\left(\frac{|h|}{H}\right)\sqrt{|h|}\overline{\rho_f}(|h|)\sum_n \tau(n)\sqrt{n}\rho_f(n)\frac{\tilde\Psi(h,n;t_f)}{\cosh(\pi t_f)}$$

for

$$\frac12 \leqslant H \leqslant \frac{eq^{1+\varepsilon}}{ab}, \tag{6.9}$$

cf. (6.2). Finally we split the $f\in\mathcal{B}(rs,\chi_0)$-sum into dyadic pieces depending on the size of t_f: namely,

$$\sum_{f\in\mathcal{B}(rs,\chi_0)} = \sum_{|t_f|<1}\dots + \sum_\tau \sum_{|t_f|\asymp\tau}\dots$$

for $\tau = 2^k$, $k\geqslant 0$ an integer. Thus typically we need to bound sums of the form

$$\sum_{rs|ae} r\mu(s) \sum_{|t_f|\asymp\tau} \sum_h G_{\overline\chi}(h;c)\rho\left(\frac{|h|}{H}\right)\sqrt{|h|}\overline{\rho_f}(|h|)\sum_n \tau(n)\sqrt{n}\rho_f(n)\frac{\tilde\Psi(h,n;t_f)}{\cosh(\pi t_f)}, \tag{6.10}$$

(plus one more sum with $\sum_{|t_f|\asymp\tau}$ replaced by $\sum_{|t_f|<1}$). Moreover, as we will see in Lemma 3 below, the contribution of the τ's greater than $q^\varepsilon\left(1+\frac{\sqrt N}{W}(\sqrt H+\sqrt R)\right)$ is negligible.

It will be useful to separate the h, n, t_f variables; we proceed by partial summation: for $j\in\mathbb{N}_0$ let

$$\Xi_j(h,n;z) := \frac{\partial^j}{\partial h^j}\frac{\partial}{\partial n}\rho\left(\frac{|h|}{H}\right)\Psi(h,n;z); \tag{6.11}$$

note that differentiation commutes with taking Bessel transforms. Then by partial summation (6.10) equals a sum of two expressions (corresponding to the signs $\pm$)

$$\sum_{rs|ae} r\mu(s)\int_0^\infty\int_0^\infty \sum_{|t_f|\asymp\tau}\frac{\tilde\Xi_1(\pm\mathfrak{h},\mathfrak{n};t_f)}{\cosh(\pi t_f)}\sum_{h\leqslant\mathfrak{h}} G_{\overline\chi}(\pm h;c)\sqrt{h}\overline{\rho_f}(h)\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\,d\mathfrak{h}\,d\mathfrak{n}, \tag{6.12}$$

but we can also suppress the partial summation with respect to h getting two expressions (corresponding to the signs $\pm$)

$$-\sum_{rs|ae} r\mu(s)\int_0^\infty \sum_{|t_f|\asymp\tau}\sum_{h\geqslant 1} G_{\overline\chi}(\pm h;c)\sqrt{h}\overline{\rho_f}(h)\frac{\tilde\Xi_0(\pm h,\mathfrak{n};t_f)}{\cosh(\pi t_f)}\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\,d\mathfrak{n}. \tag{6.13}$$

We summarize the properties of $\tilde\Xi_j(t)=\tilde\Xi_j(h,n;t)$ in the following lemma.

Lemma 3. *Let*

$$Z := \frac{q^\varepsilon R\sqrt Y}{NWae\sqrt X}\left(1+\frac{\sqrt{RN}}{W}\right)^{-1/2} \tag{6.14}$$

and

$$\tilde Z := \min\left(1,\frac{\sqrt{HN}}{W},\frac{\sqrt H}{\sqrt R}\right). \tag{6.15}$$

Then for $n \asymp N$, $|h| \asymp H$ and for any $j \in \mathbb{N}_0$ we have

$$\dot\Xi_j(t),\ \tilde\Xi_j(t),\ \check\Xi_j(t) \ll_{j,\varepsilon} \frac{Z}{1+|t|}\left(\frac{e}{H}\right)^j \tilde Z^{-2|\Im\mathfrak{m}\, t|},$$

assuming $|\Im\mathfrak{m}\, t| < 1/2$ and $t \in \mathbb{N}$ in the case of $\dot\Xi_j(t)$. Moreover, all three functions are negligible unless

$$|t| \leqslant q^\varepsilon\left(1+\frac{\sqrt{N}}{W}(\sqrt{H}+\sqrt{R})\right). \tag{6.16}$$

Proof. Let us first show that the function Ξ_j defined by (6.11) and (6.7) and supported on $z \asymp \sqrt{HN}/W$ satisfies

$$z^i\frac{\partial^i}{\partial z^i}\Xi_j(h,n;z) \ll_{i,j,\varepsilon} Z\left(\frac{e}{H}\right)^j\left(1+\frac{\sqrt{RN}}{W}\right)^i \tag{6.17}$$

for all $i,j \in \mathbb{N}_0$. To verify this we fix i and the sign of h and observe that, by the Leibniz rule for the operator $z^i(\partial^i/\partial z^i)$, the left hand side is a finite linear combination of integrals of the form (cf. (6.11) and (6.7))

$$\frac{\partial^j}{\partial h^j}\int_0^\infty\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)C(h,x,y)\,dx\,dy, \tag{6.18}$$

where we have used an obvious abstract notation and suppressed the dependence on n, z for simplicity. In particular, $A : \mathbb{R}\setminus\{0\}\times(0,\infty)\to\mathbb{C}$ is a smooth function supported on a product of compact intervals $t \asymp \pm R$, $y \asymp Y$ satisfying

$$A(t,y) \ll_\varepsilon q^\varepsilon\left(1+\frac{\sqrt{RY}}{c}\right)^{-1/2},$$

$B(t) := z^k(\partial^k/\partial z^k)\mathcal{J}^\pm\big(z\sqrt{|t|}/(4\pi)\big)$ for some $0 \leqslant k \leqslant i$ satisfying in the relevant range (cf. (6.6))

$$t^s\frac{\partial^s}{\partial t^s}B(t) \ll_{s,i,\varepsilon} q^\varepsilon\left(1+\frac{\sqrt{RN}}{W}\right)^{s+i-1/2}, \qquad z\sqrt{|t|} \asymp \frac{\sqrt{RN}}{W} \leqslant \sqrt{e}q^\varepsilon, \tag{6.19}$$

and $C : (\mathbb{R}\setminus\{0\})\times(0,\infty)\times(0,\infty)\to\mathbb{C}$ is a smooth function supported on a product of compact intervals $h \asymp H$, $x \asymp X$, $y \asymp Y$ satisfying

$$H^rX^s\frac{\partial^r}{\partial h^r}\frac{\partial^s}{\partial x^s}C(h,x,y) \ll_{r,s,i,\varepsilon} \frac{q^\varepsilon}{NW\sqrt{XY}}\left(1+\frac{\sqrt{aeXY}}{c}\right)^s. \tag{6.20}$$

Now for $j \geqslant 1$ we rewrite (6.18) as

$$\begin{aligned}&\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty\frac{\partial}{\partial h}\left\{A(h-aex,y)B\left(\frac{h-aex}{h}\right)\right\}C(h,x,y)\,dx\,dy\\ &+\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)\frac{\partial}{\partial h}C(h,x,y)\,dx\,dy.\end{aligned}$$

The inner integral in the first term equals

$$\begin{aligned}&-\frac{1}{h}\int_0^\infty A(h-aex,y)\ B_0\left(\frac{h-aex}{h}\right)C(h,x,y)\,dx\\ &+\frac{1}{ae}\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)C_0(h,x,y)\,dx,\end{aligned} \tag{6.21}$$

where

$$B_0(t) := t\frac{\partial}{\partial t}B(t), \qquad C_0(h,x,y) := \frac{\partial}{\partial x}C(h,x,y).$$

This decomposition is not obvious but follows easily by using the identities

$$\frac{\partial}{\partial h}A(h-aex,y)=-\frac{1}{ae}\frac{\partial}{\partial x}A(h-aex,y),\qquad \frac{\partial}{\partial h}B\left(\frac{h-aex}{h}\right)=-\frac{x}{h}\frac{\partial}{\partial x}B\left(\frac{h-aex}{h}\right),$$

and then integrating by parts in

$$\int_0^\infty \frac{\partial}{\partial x}\left\{A(h-aex,y)B\left(\frac{h-aex}{h}\right)\right\}C(h,x,y)\,dx.$$

From (6.19)–(6.21) we can see that (6.18) is a linear combination of 3 integrals of the form

$$\left(\frac{\sqrt{e}}{H}+\frac{1}{aeX}+\frac{\sqrt{Y}}{c\sqrt{aeX}}\right)\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty A(h-aex,y)\ B_1\left(\frac{h-aex}{h}\right)C_1(h,x,y)\,dx\,dy,$$

where A is as before; B_1 and C_1 have the same support as B and C and satisfy the same bound as in (6.19) and (6.20), respectively. By (6.4) and (6.9) we see that

$$\frac{\sqrt{e}}{H}+\frac{1}{aeX}+\frac{\sqrt{Y}}{c\sqrt{aeX}}\ll_\varepsilon q^\varepsilon\frac{e}{H}.$$

By iterating this process we can finally decompose (6.18) as a linear combination of 3^j integrals of the form

$$\left(\frac{e}{H}\right)^j\int_0^\infty\int_0^\infty A(h-aex,y)\ B_j\left(\frac{h-aex}{h}\right)C_j(h,x,y)\,dx\,dy, \tag{6.22}$$

where A is as before; B_j and C_j have the same support as B and C and satisfy the same bound as in (6.19) and (6.20), respectively. By estimating the integral pointwise we obtain (6.17) immediately.

The lemma follows now from part a) of Lemma 1, if t is real and $\sqrt{RN}/W\leqslant q^\varepsilon$. If $\sqrt{RN}/W\geqslant q^\varepsilon$ then we look closer at the first factor in the third line of (6.7). In the $\mathcal{J}^+$ case we are done by the rapid decay of the Bessel K-function. In the $\mathcal{J}^-$ case we use the asymptotic expansion of the Bessel Y-function to see that for large x,

$$\mathcal{J}^-(x)=\frac{1}{\sqrt{x}}e(2x)J_1(x)+\frac{1}{\sqrt{x}}e(-2x)J_2(x)$$

with smooth functions $J_{1,2}$ satisfying $J_{1,2}^{(j)}(x)\ll_j x^{-j}$. Now a similar argument as above together with part c) of Lemma 1 yields the proof of Lemma 3. A technical point to note here is that in this case we develop the above decomposition for $i=0$ only and then estimate the z-derivatives and the Bessel transforms inside the resulting integrals (6.22) individually. In our exposition we did not follow this path as we wanted to suppress the z-dependence for simplicity. Finally, if t is imaginary, part b) of Lemma 1 completes the proof of Lemma 3. □

We will bound separately the contribution of the τ's not exceeding a specific parameter $\mathcal{T}$ and of the τ's larger than this parameter. In the former case we shall use (6.13), in the latter (6.12).

6.2. **The case of large spectral parameter.** Using (5.5), Lemma 3 and Cauchy–Schwarz, (6.12) can be estimated from above by

$$(q^*)^{1/2}c_1q_1\sum_{rs|ae}r\sum_{d|c_2q_2}d\int_{\mathfrak{n}\asymp N}\int_{\mathfrak{h}\asymp H}\frac{Ze}{\tau H}\left(\sum_{|t_f|\asymp\tau}\frac{1}{\cosh(\pi t_f)}\left|\sum_{\substack{h\leqslant\mathfrak{h}/c_1q_1d\\(h,\frac{c_2q_2}{d})=1}}\chi(h)\sqrt{c_1q_1dh}\overline{\rho_f}(c_1q_1dh)\right|^2\right)^{\frac12}$$
$$\times\left(\sum_{|t_f|\asymp\tau}\frac{1}{\cosh(\pi t_f)}\left|\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\right|^2\right)^{\frac12}d\mathfrak{h}\ d\mathfrak{n}.$$

Decompose d into $d_2 d_2'$ such that $d_2 \mid q_2^\infty$ and $(d_2', q_2) = 1$; then for f a Hecke eigenform one has (since $(rs, q) = 1$)

$$\sqrt{c_1 q_1 dh}\overline{\rho_f}(c_1 q_1 dh) = \lambda_f(c_1 q_1 d_2)\sqrt{d_2' h}\overline{\rho_f}(d_2' h),$$

so that by the large sieve inequalities (2.14) one obtains that (6.12) is bounded by

$$\ll_\varepsilon q^\varepsilon (q^*)^{1/2}(c_1 q_1)^{1+\theta} \sum_{rs \mid ae} r \sum_{d \mid c_2 q_2} d_2' d_2^{1+\theta} \frac{Ze}{\tau H} HN \\ \times \left(\tau + \left(\frac{H}{c_1 q_1 d_2 rs}\right)^{1/2}\right)\left(\frac{H}{c_1 q_1 d}\right)^{1/2}\left(\tau + \left(\frac{N}{rs}\right)^{1/2}\right) N^{1/2}.$$

Here we clearly have the inequalities $r \leqslant rs \leqslant ae \leqslant \ell$,

$$d_2' d_2^{1+\theta}\left(\tau + \left(\frac{H}{c_1 q_1 d_2 rs}\right)^{1/2}\right)\left(\frac{H}{c_1 q_1 d}\right)^{1/2} = {d_2'}^{1/2} d_2^\theta \left(d_2^{1/2}\tau + \left(\frac{H}{c_1 q_1 rs}\right)^{1/2}\right)\left(\frac{H}{c_1 q_1}\right)^{1/2} \\ \leqslant (c_2 q_2)^{1/2+\theta}\left(\tau + \left(\frac{H}{c_1 q_1 q_2 r}\right)^{1/2}\right)\left(\frac{H}{c_1 q_1}\right)^{1/2},$$

and

$$r\left(\tau + \left(\frac{H}{c_1 q_1 q_2 r}\right)^{1/2}\right)\left(\tau + \left(\frac{N}{r}\right)^{1/2}\right) \leqslant \ell\left(\tau + \left(\frac{H}{c_1 q_1 q_2 \ell}\right)^{1/2}\right)\left(\tau + \left(\frac{N}{\ell}\right)^{1/2}\right).$$

Using these and the definition (6.14) of Z, we obtain, according to (6.4), (6.5), (6.9), (6.16), that (6.12) is bounded by

$$\begin{aligned} &\ll_\varepsilon q^\varepsilon (q^*)^{1/2}(c_1 c_2 q_1 q_2)^{1/2+\theta}\frac{\ell^{3/2} R}{\tau W}\frac{Y^{1/2}}{X^{1/2}} q^{1/2} N^{1/2}\left(\tau + \frac{q^{1/2}}{(c_1 q_1 q_2)^{1/2}}\right)\left(\tau + \frac{N^{1/2}}{\ell^{1/2}}\right) \\ &\ll_\varepsilon q^\varepsilon c^{1/2}\left(\frac{c}{q^*}\right)^\theta \ell^2 \frac{R}{\tau W} q^{1/2} N^{1/2}\left(\tau + (q^*)^{1/2}\right)\left(\tau + \frac{N^{1/2}}{\ell^{1/2}}\right). \end{aligned} \tag{6.23}$$

Let us recall that (by (6.4), (6.5), (6.9), (6.16))

$$1 \leqslant q^\varepsilon \frac{W}{N^{1/2}}\left(\frac{1}{R^{1/2}} + \frac{1}{q^{1/2}}\right), \qquad \tau \leqslant q^\varepsilon\left(1 + \frac{\ell^{1/2} q^{1/2} N^{1/2}}{W}\right), \qquad W \leqslant q^\varepsilon \ell q^{1/2};$$

we observe that the first two conditions imply that

$$\tau \leqslant q^\varepsilon \ell^{1/2}\left(1 + \left(\frac{q}{R}\right)^{1/2}\right).$$

If we assume that

$$\tau \geqslant \ell^{1/2}\mathcal{T} \qquad \text{for some} \qquad \mathcal{T} \gg q^\varepsilon,$$

then

$$R^{1/2} \ll q^{1/2+\varepsilon}\mathcal{T}^{-1} \ll q^{1/2},$$

and in particular, $(RN)^{1/2} \ll q^\varepsilon W$. Now we bound the four terms of the product

$$\frac{R}{\tau W} q^{1/2} N^{1/2}\left(\tau + (q^*)^{1/2}\right)\left(\tau + \frac{N^{1/2}}{\ell^{1/2}}\right)$$

in (6.23):

$$\tau\frac{RN^{1/2}}{W}q^{1/2} \ll_\varepsilon \tau q^\varepsilon R^{1/2}q^{1/2} \ll_\varepsilon \ell^{1/2}q^{1+\varepsilon},$$
$$\frac{RN^{1/2}}{W}(q^*)^{1/2}q^{1/2} \ll_\varepsilon q^\varepsilon R^{1/2}(q^*)^{1/2}q^{1/2} \ll_\varepsilon \frac{(q^*)^{1/2}q^{1+\varepsilon}}{\mathcal{T}},$$
$$\frac{RN}{W}\frac{q^{1/2}}{\ell^{1/2}} \ll_\varepsilon q^\varepsilon W\frac{q^{1/2}}{\ell^{1/2}} \ll_\varepsilon \ell^{1/2}q^{1+\varepsilon},$$
$$\frac{1}{\tau}\frac{RN}{W\ell^{1/2}}(q^*)^{1/2}q^{1/2} \ll_\varepsilon q^\varepsilon\frac{W}{\tau\ell^{1/2}}(q^*)^{1/2}q^{1/2} \ll_\varepsilon \frac{(q^*)^{1/2}q^{1+\varepsilon}}{\mathcal{T}}.$$

The same argument works for holomorphic forms and Eisenstein series and gives the same estimates. Therefore the total contribution of large eigenvalues to the sum (cf. (4.11))

$$\sum_{abe=\ell}\frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}}\sum_{q|c}\frac{1}{c^2}\mathcal{S}^{E,\pm}(a,b,e,c;q) \tag{6.24}$$

is bounded by

$$\ll_\varepsilon q^\varepsilon\left(\frac{\ell^2}{\mathcal{T}}\left(\frac{q^*}{q}\right)^{1/2-\theta}+\frac{\ell^{5/2}}{q^{1/2}}\left(\frac{q^*}{q}\right)^{-\theta}\right). \tag{6.25}$$

6.3. **The case of small spectral parameter.** The estimate (6.25) is useful if τ is not too small, that is, if $\mathcal{T}$ is at least some small power of q. In fact we shall later specify $\mathcal{T}$ so that $\log\mathcal{T}\asymp\log q$. In view of the preceding subsection, we suppose that

$$0\leqslant\tau\leqslant\ell^{1/2}\mathcal{T}. \tag{6.26}$$

For such small τ we use (6.13) which can be bounded by

$$\ll_\varepsilon q^\varepsilon\sum_{rs|ae}r\int_{\mathfrak{n}\asymp N}\left(\sum_{|t_f|\asymp\tau}\left|\sum_{h\geqslant1}G_{\overline{\chi}}(h;c)\sqrt{h}\overline{\rho_f}(h)\frac{\tilde{\Xi}_0(\pm h,\mathfrak{n};t_f)}{\sqrt{\cosh(\pi t_f)}}\right|^2\right)^{1/2}\left(\tau\sqrt{N}+\frac{N}{\sqrt{rs}}\right)d\mathfrak{n}, \tag{6.27}$$

using Cauchy–Schwarz and the large sieve (2.14).

For $f\in\mathcal{B}(rs,\chi_0)$ (which we recall is a Hecke eigenform), let $\mathcal{L}(f,u)$ denote the Dirichlet series

$$\mathcal{L}(f,u):=\sum_{h\geqslant1}\frac{G_{\overline{\chi}}(h;c)\sqrt{h}\rho_f(h)}{h^u}.$$

In the following we study this Dirichlet series in order to estimate the h-sum in (6.27). The Dirichlet series is absolutely convergent for $\mathfrak{Re}\,u\gg1$; by (5.5), one has

$$\begin{aligned}\mathcal{L}(f,u)&=\sum_{h_1|(rsc)^\infty}\sum_{(h_2,rsc)=1}\frac{G_{\overline{\chi}}(h_1;c)\chi(h_2)\sqrt{h_1h_2}\rho_f(h_1h_2)}{h_1^uh_2^u}\\&=L^{(rsc)}(f\otimes\chi,u)\times\overline{\chi}(c_2q_2)G_{\overline{\chi}}(1;q^*)(c_1q_1)^{1-u}\sum_{h|(rsc)^\infty}\frac{r_{c_2q_2}(h)\sqrt{c_1q_1h}\rho_f(c_1q_1h)\chi(h)}{h^u}\\&=L^{(rsc)}(f\otimes\chi,u)\times\mathcal{H}(f,u),\end{aligned}$$

say, with

$$L^{(rsc)}(f\otimes\chi,u):=\sum_{(h,rsc)=1}\frac{\lambda_f(h)\chi(h)}{h^u}$$

and

$$\mathcal{H}(f,u) := \overline{\chi}(c_2q_2)G_{\overline{\chi}}(1;q^*)(c_1q_1)^{1-u}\sum_{d|c_2q_2} d^{1-u}\chi(d)\mu\left(\frac{c_2q_2}{d}\right)\sum_{h|(rsc)^\infty}\frac{\sqrt{c_1q_1dh}\rho_f(c_1q_1dh)\chi(h)}{h^u}.$$

On the one hand,

$$L^{(rsc)}(f\otimes\chi,u) = \prod_{p|rsc}\left(1-\frac{\lambda_{\tilde{f}}(p)\chi(p)}{p^s}+\frac{\chi(p^2)}{p^{2s}}\right)\times L(\tilde{f}\otimes\chi,u)$$

where $\tilde{f}$ is the newform (of level dividing rs) underlying the Hecke eigenform f (and with the same spectral parameter t_f). Applying a subconvex bound of the form (1.2), one has

$$L^{(rsc)}(f\otimes\chi,u) \ll_\varepsilon (|u|(1+|t_f|)rsc)^\varepsilon |u|^\alpha(\tau)^\beta(rs)^\gamma(q^*)^{1/2-\delta}; \tag{6.28}$$

in particular, we remark that (1.3) is applicable if (cf. (6.26))

$$q^* \geqslant (\ell^{3/2}\mathcal{T})^4 \geqslant (rs\tau)^4. \tag{6.29}$$

On the other hand, $\mathcal{H}(f,u)$ is holomorphic for $\mathfrak{Re}\, u \geqslant 1/2$ and satisfies in this domain the uniform bound

$$\mathcal{H}(f,u) \ll (q^*c_1q_1)^{1/2}\sum_{d|c_2q_2} d^{1/2}\sum_{h|(rsc)^\infty}\frac{\sqrt{c_1q_1dh}|\rho_f(c_1q_1dh)|}{h^{1/2}}, \tag{6.30}$$

cf. [HM, (83)]. By Mellin inversion, the h-sum in (6.27) equals, without the factor $\sqrt{\cosh(\pi t_f)}$ and after replacing $f(z)$ by $\overline{f(-\bar{z})}$,

$$\frac{1}{2\pi i}\int_{(1/2)}\mathcal{L}(f,u)\left(\int_0^\infty \tilde{\Xi}_0(\pm x,\mathfrak{n};t_f)\,x^{u-1}\,dx\right)du.$$

By partial integration and Lemma 3 we see

$$\int_0^\infty \tilde{\Xi}_0(x,\mathfrak{n};t_f)\,x^{u-1}\,dx \ll_\varepsilon \frac{Z\sqrt{H}}{1+\tau}\left(\frac{e}{|u|}\right)^\nu \tilde{Z}^{-2|\Im\mathfrak{m}\, t_f|} \tag{6.31}$$

on $\mathfrak{Re}\, u = 1/2$, for any $\nu \geqslant 0$ (at first for integer ν, but then by convexity also for real ν). We choose $\nu := \alpha+1+\varepsilon$ in order to ensure absolute convergence of the u-integral. Using Cauchy–Schwarz and (2.15), we see that

$$\begin{aligned}\left(\sum_{|t_f|\asymp\tau}\left|\sum_{h|(rsc)^\infty}\frac{\sqrt{c_1q_1dh}|\rho_f(c_1q_1dh)|}{h^{1/2}\sqrt{\cosh(\pi t_f)}}\right|^2\right)^{\frac12} &\ll_\varepsilon (rsc)^\varepsilon\sum_{h|(rsc)^\infty}\left(\frac{1}{h^{1-\varepsilon}}\sum_{|t_f|\asymp\tau}\frac{c_1q_1dh|\rho_f(c_1q_1dh)|^2}{\cosh(\pi t_f)}\right)^{\frac12}\\ &\ll_\varepsilon c^\varepsilon(c_1q_1d)^\theta(1+\tau),\end{aligned} \tag{6.32}$$

where $\theta = 7/64 < 1/2$ (cf. (2.4)–(2.5)). Collecting the estimates (6.28), (6.30), (6.31), (6.32), we can bound (6.27) by

$$\begin{aligned}&\ll_\varepsilon q^\varepsilon\sum_{rs|ae} re^{\alpha+1}(\tau)^\beta(rs)^\gamma(q^*)^{1-\delta}\sum_{d|c_2q_2}(c_1q_1d)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{rs}}\right)\tilde{Z}^{-2\theta_0}\\ &\ll_\varepsilon c^{2\varepsilon}(\tau)^\beta\ell^{2+\alpha+\gamma}(q^*)^{1-\delta}\left(\frac{c}{q^*}\right)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{\ell}}\right)\tilde{Z}^{-2\theta_0},\end{aligned} \tag{6.33}$$

where $\theta_0 = 0$ if $\tau \geqslant 1$ and $\theta_0 = \theta$ if $\tau \leqslant 1$.

If $\tau \geqslant 1$, $\tilde{Z}^{-2\theta_0} = 1$ and we use the bound (6.26) to obtain that (6.33) is at most

$$c^{2\varepsilon}\mathcal{T}^\beta\ell^{2+\alpha+\beta/2+\gamma}(q^*)^{1-\delta}\left(\frac{c}{q^*}\right)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{\ell}}\right). \tag{6.34}$$

We deal now with the sum (6.27) where the summation $\sum_{|t_f|\asymp\tau}$ is replaced by $\sum_{|t_f|<1}$: we recall that $\tilde{Z}$ depends on H according to (6.15), so that (6.33) is an increasing function of H. Thus we estimate (6.33) from above using (6.9). But then, together with (6.4), we see that $\tilde{Z} \geqslant \ell^{-1/2}q^{-\varepsilon}$ so that $\tilde{Z}^{-2\theta_0} \leqslant q^{\varepsilon}\ell^{\theta}$; in that case however, there is no factor $(\mathcal{T}\sqrt{\ell})^{\beta}$. Since $\beta \geqslant 3/8 > 2\theta$, the contribution of $|t_f| < 1$ is dominated by (6.34).

Using (6.14), and the bound (cf. (6.4) and (6.9))

$$\frac{\sqrt{Y}}{ae\sqrt{X}}\sqrt{H} \ll_{\varepsilon} q^{1/2+\varepsilon},$$

we are left with

$$c^{2\varepsilon}\mathcal{T}^{\beta}\ell^{2+\alpha+\beta/2+\gamma}(q^*)^{1-\delta}q^{1/2}\left(\frac{c}{q^*}\right)^{1/2+\theta}\frac{RN^{1/2}}{W}\left(1+\frac{\sqrt{RN}}{W}\right)^{-1/2}\left(\mathcal{T}\sqrt{\ell}+\frac{\sqrt{N}}{\sqrt{\ell}}\right),$$

subject to

$$\frac{\sqrt{RN}}{W} \leqslant \sqrt{\ell}q^{\varepsilon}, \qquad R \leqslant \ell q^{1+\varepsilon}, \qquad W \leqslant \ell q^{1/2+\varepsilon}.$$

Averaging over $c \equiv 0 \pmod q$, we see that the total contribution of small eigenvalues to (6.24) is at most

$$\begin{aligned}
&\ll_{\varepsilon} q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+2}\mathcal{T}^{\beta}\frac{(q^*)^{1/2-\theta-\delta}}{q^{1-\theta}}\frac{R^{\frac{3}{4}}N^{\frac{1}{4}}}{W^{\frac{1}{2}}}\left(\mathcal{T}\sqrt{\ell}+\frac{\sqrt{N}}{\sqrt{\ell}}\right)\\
&\ll_{\varepsilon} q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+2}\mathcal{T}^{\beta}\frac{(q^*)^{1/2-\theta-\delta}}{q^{1-\theta}}\left(R^{\frac{1}{2}}\mathcal{T}\ell^{\frac{3}{4}}+W\ell^{\frac{1}{4}}\right),\\
&\ll_{\varepsilon} q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}}\mathcal{T}^{\beta+1}\left(\frac{q^*}{q}\right)^{1/2-\theta}(q^*)^{-\delta}. \qquad (6.35)
\end{aligned}$$

The same bound holds for holomorphic cusp forms. The case of Eisenstein series is somewhat different at least when they are parametrized by the cusps for their Fourier coefficients are not multiplicative anymore. Instead we proceed as in [Mi, HM] and calculate the coefficients directly. Unfolding the Gauss sum leads for each cusp $\mathfrak{a} = \frac{v}{w}$, $w \mid rs$, to the normalized series

$$\sum_{h}\frac{\chi(h)\sqrt{gh}\overline{\rho_{\mathfrak{a}}}(1/2+it,gh)}{h^u\sqrt{\cosh(\pi t)}}, \qquad (6.36)$$

where $g := c_1q_1dd'$ and $dd' \mid c_2q_2$. By the computation of [HM, Section 5.4.2] this series can be written in terms of products of two Dirichlet L-functions $L(\chi\overline{\varphi}, u-it)L(\chi\varphi, u+it)$ for certain characters φ having conductor $\leqslant \left(w, \frac{rs}{w}\right)$, times a holomorphic function in $\mathfrak{Re}\, u \geqslant 1/2$ that is bounded on $\mathfrak{Re}\, u = 1/2$ by

$$\ll_{\varepsilon} (grs)^{\varepsilon}\,(g,w)\left(w,\frac{rs}{w}\right)^{1/2}(rs)^{-1/2}.$$

Here we used that $(rs,q)=1$. In particular, the function defined by (6.36) can be holomorphically continued to $\mathfrak{Re}\, u \geqslant 1/2$ and on $\mathfrak{Re}\, u = 1/2$ it is bounded by

$$\ll_{\varepsilon} (q(1+|t|)|u|)^{\varepsilon}\,(|u|+|t|)^{3/8}(g,rs)\left(w,\frac{rs}{w}\right)^{7/8}(rs)^{-1/2}q^{3/8},$$

according to Heath-Brown's hybrid bound [HB] for Dirichlet L-functions. Summing over all cusps of $\Gamma_0(rs)$ and noting that

$$\sum_{w\mid rs}\varphi\left(w,\frac{rs}{w}\right)\left(w,\frac{rs}{w}\right)^{7/8}(rs)^{-1/2} \ll_{\varepsilon} (rs)^{7/16+\varepsilon},$$

we obtain a bound of at least the same quality as in the case of Maaß cusp forms if we assume $\alpha, \beta \geqslant 3/8$, $\gamma \geqslant 7/16$, $\delta \leqslant 1/8$. Then we proceed analogously.

7. Concluding the Proof of Theorem 2

Collecting (4.2), (4.8), (4.16), (5.9), (6.25) and (6.35), we obtain that

$$\begin{aligned} k^{-18} \left|\mathcal{Q}_k^{\mathrm{holo}}(\ell)\right| + |\mathcal{Q}(\ell)| &\\ \ll_{s,t_0,\varepsilon} q^{\varepsilon} &\left(\frac{1}{\ell^{1/2}} + \frac{\ell^2}{\mathcal{T}} \left(\frac{q^*}{q}\right)^{1/2-\theta} + \frac{\ell^{5/2}}{q^{1/2}} \left(\frac{q^*}{q}\right)^{-\theta} + \ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}} \mathcal{T}^{\beta+1} \left(\frac{q^*}{q}\right)^{1/2-\theta} (q^*)^{-\delta} \right) \\ \ll_{s,t_0,\varepsilon} q^{\varepsilon} &\left(\frac{1}{\ell^{1/2}} + \frac{\ell^2}{\mathcal{T}} \left(\frac{q^*}{q}\right)^{1/2-\theta} + \ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}} \mathcal{T}^{\beta+1} (q^*)^{-\delta} \left(\frac{q^*}{q}\right)^{1/2-\theta} \right) \end{aligned} \tag{7.1}$$

(in the first inequality above the last term is always larger than the third one).

Set $q^* = q^\eta$ with $\eta \in [0,1]$. If η is small (to be determined in a moment) we choose $\mathcal{T} := q^\varepsilon \sqrt{\ell}$ and apply the convexity bound (cf. (1.2)) with

$$\alpha = \frac{1}{2}, \qquad \beta = \frac{1}{2}, \qquad \gamma = \frac{1}{4}, \qquad \delta = 0,$$

and so we arrive at we arrive at (3.20) with

$$c_1 := 5 \qquad \text{and} \qquad c_2 := (1-\eta)\left(\frac{1}{2} - \theta\right).$$

Substituting the expressions for c_1 and c_2 into (3.23) we obtain

$$L(f_0, s) \ll_{s,t_0,\varepsilon} q^{\frac{1}{4} - \frac{(1-\eta)(1-2\theta)}{168} + \varepsilon}. \tag{7.2}$$

If η is large, we use the exponents provided by (1.3),

$$\alpha := \frac{1}{2}, \qquad \beta := 3, \qquad \gamma := \frac{1}{4}, \qquad \delta := \frac{1}{8},$$

assuming that (6.29) holds. Equating the second and third terms of (7.1) we choose

$$\mathcal{T} := (q^*)^{\frac{\delta}{\beta+2}+\varepsilon} \ell^{-\frac{\alpha+\beta/2+\gamma+5/4}{\beta+2}},$$

provided

$$q^{\eta\delta-\varepsilon} > \ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{5}{4}} \tag{7.3}$$

(so that $\log \mathcal{T} \asymp \log q$), and provided

$$q^{\eta(\beta+2-4\delta)-\varepsilon} > \ell^{-4\alpha+4\beta-4\gamma+7} \tag{7.4}$$

(in order to satisfy (6.29)). Under these assumptions we obtain a total error term of

$$\ll_\varepsilon q^\varepsilon \left(\frac{1}{\ell^{1/2}} + \frac{\ell^{\frac{\alpha+5\beta/2+\gamma+21/4}{\beta+2}}}{q^{\eta\frac{\delta}{\beta+2} + (1-\eta)\left(\frac{1}{2}-\theta\right)}} \right) \ll_\varepsilon q^\varepsilon \left(\frac{1}{\ell^{1/2}} + \frac{\ell^{\frac{\alpha+5\beta/2+\gamma+21/4}{\beta+2}}}{q^{\frac{\delta}{\beta+2}}} \right),$$

since $\frac{1}{2} - \theta \geqslant \frac{\delta}{\beta+2}$ for any $\beta \geqslant 0$ and any $\delta \in [0, 1/2]$. Hence we arrive at (3.20) with

$$c_1 := \frac{\alpha + 5\beta/2 + \gamma + 21/4}{\beta+2} \qquad \text{and} \qquad c_2 := \frac{\delta}{\beta+2}.$$

We choose L as in (3.22):

$$L := q^{c_2/(2c_1+1/2)}.$$

In (3.21) we apply (3.20) for $\ell \leqslant L^2$, and it is easily checked that (7.3) and (7.4) are satisfied as long as $\eta \geqslant 14/59$. Substituting the expressions for c_1 and c_2 into (3.23) we obtain

$$L(f_0, s) \ll_{s,t_0,\varepsilon} q^{\frac{1}{4} - \frac{\delta}{16\alpha+44\beta+16\gamma+92} + \varepsilon} \ll_{s,t_0} q^{\frac{1}{4} - \frac{1}{1889}} \tag{7.5}$$

for $\eta \geqslant 14/59$ while for $\eta \leqslant 14/59$ the bound (7.2) is stronger. This concludes the proof of Theorem 2.

References

[BH] V. Blomer, G. Harcos, *Hybrid bounds for twisted L-functions*, J. Reine Angew. Math., to appear 2

[BHM] V. Blomer, G. Harcos, P. Michel, *A Burgess-like subconvex bound for twisted L-functions (with Appendix 2 by Z. Mao)*, Forum Math. **19** (2007), 61–105 2, 10

[Bu] D. A. Burgess, *On character sums and L-series. II*, Proc. Lond. Math. Soc. **13** (1963), 524–536 3

[By] V. A. Bykovskiĭ, *A trace formula for the scalar product of Hecke series and its applications*, translated in J. Math. Sci (New York) **89** (1998), 915–932 3, 5

[DI] J.-M. Deshouillers, H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. Math. **70** (1982), 219–288 9, 10

[Du] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. 92 (1988), 73–90 3

[DFI1] W. Duke, J. Friedlander, H. Iwaniec, *A quadratic divisor problem*, Invent. Math. **115** (1994), 209–217 4

[DFI2] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions. II*, Invent. Math. **115** (1994), 219–239 5

[DFI3] W. Duke, J. Friedlander, H. Iwaniec, *Representations by the determinant and mean values of L-functions*, In: Sieve methods, exponential sums, and their applications in number theory (Cardiff 1995), 109–115, London Math. Soc. Lecture Note Ser. 237, Cambridge Univ. Press, Cambridge, 1997 4

[DFI4] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions. III*, Invent. Math. **143** (2001), 221–248 5

[DFI5] W. Duke, J. Friedlander, H. Iwaniec, *The subconvexity problem for Artin L-functions*, Invent. Math. **149** (2002), 489–577 1, 2, 3, 4, 5, 7, 13, 14

[ELMV] M. Einsiedler, E. Lindenstrauss, P. Michel, A. Venkatesh, *Distribution of periodic torus orbits and Duke's theorem for cubic fields*, submitted 3

[Fr] J. Friedlander, *Bounds for L-functions*, Proc. Int. Congr. Math. (Zürich 1994) Vol. II, 363–373, Birkhäuser, Basel, 1995 1

[GJ] S. Gelbart, H. Jacquet, *Forms on GL_2 from the analytic point of view*, In: Automorphic forms, representations, and L-functions (A. Borel, W. Casselman eds.), Part 1, Proc. Sympos. Pure Math. **33** (1979), 213–251 6

[GR] I. S. Gradshteyn, I. M. Ryzhik, *Tables of integrals, series, and products*, 5th edition, Academic Press, New York, 1994 10, 11, 21

[Ha] G. Harcos, *Uniform approximate functional equation for principal L-functions*, Int. Math. Res. Not. **2002**, 923–932.; *Erratum*, ibid. **2004**, 659–660 13

[HM] G. Harcos, P. Michel, *The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points. II*, Invent. Math. **163** (2006), 581–655 2, 3, 7, 9, 18, 33, 34

[HB] D. R. Heath-Brown, *Hybrid bounds for Dirichlet L-functions. II*, Quart. J. Math. Oxford Ser. (2), **31** (1980), 157–167 34

[Iw] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd edition, Graduate Studies in Mathematics 53, American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002 5, 8

[IK] H. Iwaniec, E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications 53, American Mathematical Society, Providence, RI, 2004 8

[IS] H. Iwaniec, P. Sarnak, *Perspectives in the analytic theory of L-functions*, Geom. Funct. Anal. 2000, Special Volume, Part II, 705–741 1

[Ju] M. Jutila, *Convolutions of Fourier coefficients of cusp forms*, Publ. Inst. Math. (Beograd) (N.S.) **65(79)** (1999), 31–51 10

[KS] H. Kim, *Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2 (with Appendix 1 by D. Ramakrishnan and Appendix 2 by H. Kim and P. Sarnak)*, J. Amer. Math. Soc. **16** (2003), 139–183 6

[KMV] E. Kowalski, P. Michel, J. VanderKam, *Mollification of the fourth moment of automorphic L-functions and arithmetic applications*, Invent. Math. **142** (2000), 95–151 3, 5, 11, 19, 20, 21

[Me] T. Meurman, *On the binary additive divisor problem*, In: Number theory (Turku 1999), 223–246, de Gruyter, Berlin, 2001 4, 11

[Mi] P. Michel, *The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. **160** (2004), 185–236 2, 3, 4, 34

[MV] P. Michel, A. Venkatesh, *Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik*, Proc. Int. Congr. Math. (Madrid 2006), Vol. II, 421–457, Eur. Math. Soc., Zürich, 2006 1

[Pr] N. V. Proskurin, *On the general Kloosterman sums*, translated in J. Math. Sci. (New York) **129** (2005), 3874–3889 7, 8

Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, Ontario, Canada, M5S 2E4
E-mail address: vblomer@math.toronto.edu

The University of Texas at Austin, Mathematics Department, 1 University Station C1200, Austin, TX 78712-0257, USA
Current address: Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, Budapest H-1364, Hungary
E-mail address: gharcos@renyi.hu

I3M, UMR CNRS 5149, Université Montpellier II CC 051, 34095 Montpellier Cedex 05, France
E-mail address: michel@math.univ-montp2.fr