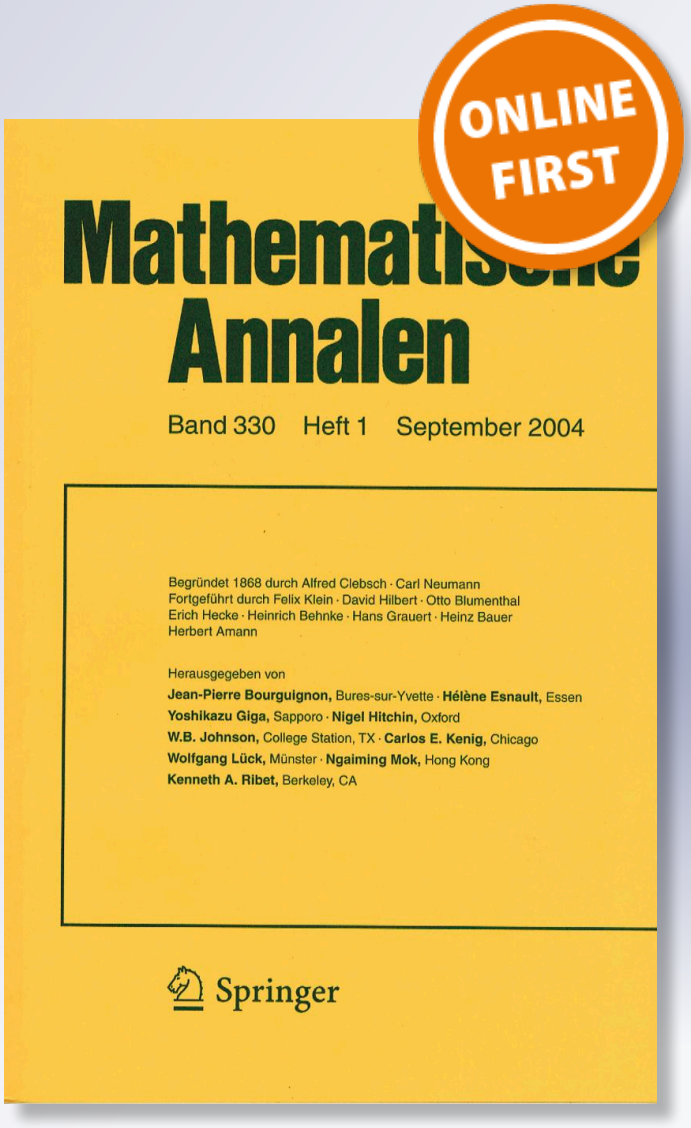
A generalized cubic moment and the Petersson formula for newforms

Ian Petrow & Matthew P. Young

Mathematische Annalen

ISSN 0025-5831

Math. Ann.
DOI 10.1007/s00208-018-1745-1



Springer

Mathematische Annalen
https://doi.org/10.1007/s00208-018-1745-1

Mathematische Annalen

A generalized cubic moment and the Petersson formula for newforms

Ian Petrow[1] · Matthew P. Young[2]

Received: 12 June 2018 / Revised: 6 August 2018

Abstract
Using a cubic moment, we prove a Weyl-type subconvexity bound for the quadratic twists of a holomorphic newform of square-free level, trivial nebentypus, and arbitrary even weight. This generalizes work of Conrey and Iwaniec in that the newform that is being twisted may have arbitrary square-free level, and also that the quadratic character may have even conductor. One of the new tools developed in this paper is a more general Petersson formula for newforms of square-free level.

Mathematics Subject Classification 11F11 · 11F37 · 11F66 · 11M99

1 Introduction

1.1 Cubic moments

Let χ_q be a real, primitive character of conductor q and $\tilde{q} = \mathrm{rad}(q)$ its square-free kernel. Let $H_\kappa^*(N)$ be the set of Hecke-normalized holomorphic newforms for $\Gamma_0(N)$, of weight κ, and trivial central character. Our main result is

Theorem 1 *For any square-free r with $(r, q) = 1$ we have*

$$\sum_{\substack{f \in H_\kappa^*(rq') \\ q' | \tilde{q}}} L(1/2, f \otimes \chi_q)^3 \ll_{\kappa,\varepsilon} (qr)^{1+\varepsilon}. \tag{1}$$

The estimate holds for any even $\kappa \geq 2$, and depends polynomially on κ.

Communicated by Kannan Soundararajan.

✉ Ian Petrow
ian.petrow@math.ethz.ch

Matthew P. Young
myoung@math.tamu.edu

[1] Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zurich, Switzerland

[2] Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA

Published online: 21 August 2018

Corollary 1 *For any holomorphic newform f of square-free level s and χ_q any real primitive character of conductor q we have*

$$L(1/2, f \otimes \chi_q) \ll \left(\frac{sq}{(s,q)}\right)^{1/3+\varepsilon}. \tag{2}$$

Remark The conductor of $L(1/2, f \otimes \chi_q)$ is $sq^2/(s,q)$. Therefore, the bound (2) is a Weyl-type subconvexity bound in q-aspect, but does not reach the convexity bound in the s-aspect. Note that the corollary holds without a relatively prime hypothesis on s, q.

Corollary 1 gives a non-trivial bound when the root number $\epsilon_{f\otimes\chi_q} = +1$ (since otherwise $L(1/2, f \otimes \chi_q) = 0$). See Sect. 8.1, specifically Eq. (64) for a concrete formula for the root number.

Our work here is a generalization of the cubic moment studied by Conrey and Iwaniec [14], who obtained, in our notation, the case $r = 1$, $\kappa \geq 12$, and q odd. The extension of their work to $\kappa \geq 2$ was obtained by the first-named author [40]. It may be somewhat surprising that, prior to Corollary 1, a Weyl-type subconvex bound in the q-aspect was previously not known for any values of r besides 1 (nor for even q). The case $r = 1$ has some pleasant simplifications; for one, the conductor of $L(1/2, f \otimes \chi_q)$ is q^2 for all f of level dividing q. Furthermore, the nth Fourier coefficient of $f \otimes \chi_q$ vanishes unless $(n, q) = 1$. For these reasons, Conrey and Iwaniec could use a formula of Iwaniec, Luo, and Sarnak [26], who proved a Petersson formula that is applicable to (1) with $r = 1$. The case $r \neq 1$ lacks these simplifications, so in order to approach the proof of Theorem 1 we developed a more general form of the Petersson formula that is applicable to (1) with any square-free r. This formula, which is of independent interest, is described in Sect. 1.3.

Corollary 1 improves on a hybrid subconvexity result of Blomer and Harcos [8, Theorem 2'], which holds more generally for f of arbitrary level and nebentype character. In our notation (and assuming $(q, r) = 1$) the result of Blomer and Harcos takes the form

$$L(1/2, f \otimes \chi_q) \ll (r^{1/4}q^{3/8} + r^{1/2}q^{1/4})(rq)^{\varepsilon}. \tag{3}$$

One may check that Corollary 1 is superior to (3), except in the range $r \asymp q^{1/2+o(1)}$ where all the bounds are equalized. This result of Blomer and Harcos is more general in that χ_q can be replaced by an arbitrary primitive Dirichlet character, f may be a Maass form, and it is not restricted to the central point. In addition, the Blomer-Harcos bound proceeds by bounding an amplified second moment, and is Burgess-quality in the q-aspect for r fixed. If q is fixed and r is large, then the cubic moment is not the appropriate moment to use, and both Corollary 1 and (3) are weaker than the convexity bound [specifically, (2) is superior to the convexity bound of $(rq^2)^{1/4+\varepsilon}$ for $r \ll q^{2-\varepsilon}$].

The work of [14] treats both holomorphic forms and Maass forms, with similar proofs. Provided one generalizes our newform Petersson formula to the setting of the Bruggeman–Kuznetsov formula, then our methods should carry over to the Maass case, as in [14]. Note added May 31, 2018: the Bruggeman–Kuznetsov formula for newforms has now appeared in [46].

The type of sum appearing in Theorem 1 may look somewhat unusual, but it is very important for the proof. It is crucial in [14] that, after applying the Petersson formula, the moduli of the Kloosterman sums are all divisible by q. The form of (1) is chosen to group together the terms with $q' \mid q$ to give a sum of Kloosterman sums with $c \equiv 0 \pmod q$. As a rough sketch of what this means, and why it is important, one may consider the case of prime level q. Very roughly, one naively expects the Petersson formula to say

$$\sum_{f \text{ level } 1} \frac{a_f(m)a_f(n)}{\langle f, f\rangle} + \sum_{f \text{ new of level } q} \frac{a_f(m)a_f(n)}{\langle f, f\rangle}$$
$$\leftrightarrow \delta_{m,n} + 2\pi i^{-\kappa} \sum_{c \equiv 0 (\text{mod } q)} \frac{S(m, n; c)}{c}. \tag{4}$$

This is not quite correct because there are other types of oldforms not appearing on the left hand side, but that does not affect the broader thrust of this discussion (the reader interested in the correct version of this formula will find abundant discussion throughout this paper!). Meanwhile, the sum over f of level 1 has a Petersson formula in which *all* $c \geq 1$ appear. Thus, by rearranging these expressions, we see that a newform formula for f of level q should have all $c \geq 1$ present. With the cubic moment, one also has a factor $\chi_q(mn)$, and one wishes to apply Poisson in these variables. The total modulus of $\chi_q(mn)S(m, n; c)$ is $[q, c]$ which for $q \mid c$ is still c, but if $(q, c) = 1$ it is qc which is much larger. In this latter case, Poisson summation is practically ineffectual.

Our proof of Theorem 1 in fact shows a stronger asymptotic result of the form

$$\sum_{\substack{f \in H_\kappa^*(rq') \\ q' \mid \tilde{q}}} \omega_f L(1/2, f \otimes \chi_q)^3 = R_{r,q} + O((qr)^\varepsilon (r^{-1/2} + q^{-1/2} r^{-1/4})), \tag{5}$$

where ω_f are certain positive weights satisfying $\omega_f = (qr)^{-1+o(1)}$ and $R_{r,q}$ is a complicated main term arising from a residue calculation (see Sect. 8.4 for details). The error term here is seen to be $o(1)$ provided $r \gg q^\delta$ for some fixed $\delta > 0$. Conrey and Iwaniec [14] express interest in finding the asymptotic of the cubic moment in their case $r = 1$; it is perhaps surprising that deforming the problem slightly in the r-aspect allows us to solve this problem in a hybrid range. In light of (5), perhaps it is possible to amplify the moment in the r-aspect, and thereby improve the exponent of s in (2).

1.2 Arithmetical applications of the cubic moment

The bound from Corollary 1 implies a bound on the Fourier coefficients of half-integral weight cusp forms, as we now describe. Suppose that $g(z) = \sum_n c(n)e(nz)$ is a weight $\frac{\kappa+1}{2}$ Hecke eigenform of nebentypus ψ and level $4r$ where r is odd and square-free, and κ is even. The Shimura correspondence links g to a form f of weight κ, level $2r$,

and nebentypus ψ^2 , and Waldspurger's formula gives under some local conditions (see [43, Théorème 1]) that $c(|D|)^2 = c_0|D|^{\frac{\kappa-1}{2}} L(1/2, f \otimes \psi^{-1}\chi_D)$ where D is a fundamental discriminant, and c_0 is some constant of proportionality depending on g. Note $f \otimes \psi^{-1}$ has trivial nebentypus. Since $2r$ is square-free, Corollary 1 applies, and we deduce:

Theorem 2 *Assume that ψ is a real character. With notation as above, we have*

$$c(|D|) \ll_g |D|^{\frac{\kappa-1}{4}+\frac{1}{6}+\varepsilon}. \tag{6}$$

Theorem 2 has applications to the representation problem for ternary quadratic forms which has been studied by a number of authors, including [6,7,17,18,27] from which we have drawn some of the following background material. Suppose that Q is a positive ternary quadratic form with associated theta function θ_Q of level dividing $4N$ with N odd and square-free. For instance, any diagonal form $ax^2 + by^2 + cz^2$ with abc odd and square-free satisfies these conditions. Then $\theta_Q = E + U + S$ where E is a linear combination of Eisenstein series, U is a linear combination of unary theta functions, and S is a linear combination of Hecke cusp forms. According to this decomposition, write $r_Q(n) = c_E(n) + c_U(n) + c_S(n)$ where $r_Q(n)$ is the number of representations of n by Q, and $c_*(n)$ is the nth Fourier coefficient of $* = E, U, S$. For ease of exposition, suppose that n is square-free and coprime to the level, which implies $c_U(n) = 0$. If n is locally represented everywhere by Q, then $c_E(n) \gg_Q n^{1/2-\varepsilon}$. Theorem 2 implies $c_S(n) \ll_Q n^{5/12+\varepsilon}$, which is an improvement over that derived from the Burgess-quality subconvex bound of [8].

For some more advanced questions, one may desire to explicate the dependence on g in Theorem 2. Blomer [7] remarks that in general this is difficult, and that Mao [9, Appendix 2] has done this but at the expense of relating the Fourier coefficients to twisted L-values of an auxiliary form $f \otimes \psi^{-1}\chi_{-4}^{\kappa/2}$, which is of level dividing $16r^2$. Our results here then may not apply to this auxiliary form .

However, if g is in Kohnen's plus space, then the constant of proportionality is given explicitly by the Kohnen–Zagier formula, and our results apply, as we now explain. We gather some notation from Kohnen's paper [30], paying careful attention to normalizations. Let g be as defined in this section, in Kohnen's plus space, and write $f(z) = \sum_{n=1}^{\infty} n^{\frac{\kappa-1}{2}} \lambda_f(n) e(nz)$ with $\lambda_f(1) = 1$. Define the Petersson inner product by

$$\langle f, f\rangle_{\text{Kohnen}} = \frac{1}{[\Gamma_0(1) : \Gamma_0(r)]} \int_{\Gamma_0(r)\backslash\mathbb{H}} y^{\kappa} |f(z)|^2 \frac{dxdy}{y^2}.$$

(For the rest of the paper we will mainly use a different normalization of the inner product.) Using [26, Lemma 2.5] [22,23], we have $\langle f, f\rangle_{\text{Kohnen}} = r^{o(1)}$. Let D be a fundamental discriminant with $(-1)^{\kappa/2}D > 0$, coprime to r. By [30, Corollary 1], we have

$$\frac{|c(|D|)|^2}{\langle g, g\rangle_{\text{Kohnen}}} = 2^{\nu(r)} \frac{\left(\frac{\kappa}{2}-1\right)!}{\pi^{\kappa/2}} |D|^{\frac{\kappa-1}{2}} \frac{L(1/2, f \otimes \chi_D)}{\langle f, f\rangle_{\text{Kohnen}}},$$

under the assumption $\chi_D(p) = \eta_p(f)$ for all $p \mid r$ (here $\eta_p(f)$ is the eigenvalue of the Atkin–Lehner operator). If $\chi_D(p) = -\eta_p(f)$ for some $p \mid r$ then $c(|D|) = 0$ while the right hand side may not vanish. As an aside, we mention that Baruch and Mao [5] have generalized the Kohnen–Zagier/Waldspurger formula by removing these conditions on D, relating the central value to a Fourier coefficient of a different half-integral weight cusp form. By Corollary 1, we have

$$\frac{|c(q)|^2}{\langle g, g\rangle_{\text{Kohnen}}} \ll_\kappa r^{\frac{1}{3}+\varepsilon} q^{\frac{\kappa-1}{2}+\frac{1}{3}+\varepsilon}. \tag{7}$$

It is also natural to inquire into the normalization of the form g. There is a slight difficulty here in that we cannot scale g by taking $c(1) = 1$, since $c(1)$ may vanish. There exists a D_0, polynomially bounded in r, so that $L(1/2, f \otimes \chi_{D_0}) \gg r^{o(1)}$ (e.g., see [21]). Then we may choose the constant of normalization so that $|c(|D_0|)|^2 = |D_0|^{\frac{\kappa-1}{2}}$. Then with this normalization, $\langle g, g\rangle_{\text{Kohnen}} \ll r^{o(1)}$, and hence

$$|c(q)| \ll_{\kappa,\varepsilon} r^{\frac{1}{6}+\varepsilon} q^{\frac{\kappa-1}{4}+\frac{1}{6}+\varepsilon}. \tag{8}$$

Theorem 1 itself can be used to improve many exponents in the results of [34]. In particular, we improve the rate of equidistribution of the reductions of CM elliptic curves (see [34] for a full description of this arithmetical problem). For brevity, we shall not repeat any material from [34], but will instead indicate which exponents may be improved. The bound $q^{1/8+\varepsilon} D^{7/16+\varepsilon}$ in [34, (1.5)] may be replaced by $q^\varepsilon D^{5/12+\varepsilon}$. In [34, Corollary 1.3], the bound $D \gg q^{18+\varepsilon}$ may be replaced by $D \gg q^{12+\varepsilon}$. In [34, (1.10), (1.12)], the bound $q^{7/8} D^{7/16}$ may be replaced by $q^{3/4} D^{5/12}$. All these changes result from a use of Theorem 1 to bound M_2 defined by [34, (4.7), (3.1)] with

$$M_2 \ll \frac{D^{1/2+\varepsilon}}{q^{1/3-\varepsilon}} \left(\sum_{f \in H_2^*(q)} L(1/2, f \otimes \chi_D)^3 \right)^{1/3} \ll q^\varepsilon D^{5/6+\varepsilon}. \tag{9}$$

If one can generalize Theorem 1 (and hence Corollary 1) to allow f to be a Hecke-Maass cusp form, then there are additional applications. This is the setting required for equidistribution of integral points on ellipsoids [16]. The various exponents appearing in [33] would be updated similarly to the improvements to [34] described in the previous paragraph. As another example in this vein, Folsom and Masri [19,35] have improved the error term in the asymptotic formula for the partition function which requires subconvexity for quadratic twists of a cusp form of level 6; the previous bounds of [14] do not apply, and so the methods developed in this paper pave the way for further improvements.

The second-named author [45] generalized the method of [14] allowing for large weights (or spectral parameters, in the Maass case) giving a Weyl-type hybrid subconvexity bound. This had applications to equidistribution problems on shrinking sets. For simplicity, in this paper we have kept the weight κ fixed but it seems likely that the methods of [45] could be combined with those in this paper to allow κ to vary.

1.3 Petersson formula for newforms

We begin with an expanded discussion on why a newform Petersson formula is relevant for Theorem 1.

One encounters a significant difficulty when attempting to generalize the method of [14] to allow level structure of the base form, as we describe. One begins by using an approximate functional equation of the L-function $L(1/2, f \otimes \chi_q)$, which has conductor rq^2 when f is a newform of level rq' with $(r, q) = 1$ and $q' \mid q$. Next one would wish to apply the Petersson formula to average over an orthogonal basis of cusp forms. The problem is that this basis consists of oldforms as well as newforms, which causes a variety of problems. Firstly, it is not clear what Dirichlet series to attach to $f \otimes \chi_q$ when f is an oldform. One could take this to mean that if f is induced from a newform f^* of lower level, then we take $L(1/2, f^* \otimes \chi_q)$. However, with this definition the conductor of this L-function may be a divisor of rq^2, in which case there is some dependence on the level of f^* in the approximate functional equation. The classical Petersson formula is unable to distinguish between these forms.

It is plausible that there is some trick that lets one set up the problem to prove Theorem 1 using the classical Petersson formula, but the authors are not aware of one (if the moment was an even power, this would be easy because of positivity; the fact that the moment is an odd power in this application makes this more difficult).

The robust solution is to prove a Petersson formula for the newforms only, similarly to the existence of averaging formulas for primitive Dirichlet characters of a given modulus (see [25, (3.8)]). Iwaniec, Luo, and Sarnak have proven a Petersson formula for newforms of square-free level [26, Proposition 2.8], but with some coprimality conditions on the level and the Fourier coefficients of the modular forms, which in our application are crucial to avoid. When working with 1-level density of zeros of L-functions, it is easy to ensure coprimality because the log derivative of an L-function is a sum over prime powers. However, the L-function itself is not so easily treated, because altering a single Euler factor will ruin the functional equation. For this reason, we have generalized the [26] formula to hold with square-free level and arbitrary Fourier coefficients.

Suppose that N is a positive integer, and let $B_\kappa(N)$ denote an orthogonal basis for the space of weight κ cusp forms for $\Gamma_0(N)$. For $f \in B_\kappa(N)$, write $f(z) = \sum_{n=1}^{\infty} a_f(n) e(nz)$, and $a_f(n) = \lambda_f(n) n^{\frac{\kappa-1}{2}}$. Let

$$\Delta_N(m, n) = c_\kappa \sum_{f \in B_\kappa(N)} \frac{\overline{\lambda_f(m)} \lambda_f(n)}{\langle f, f \rangle_N}, \quad \text{where} \quad c_\kappa = \frac{\Gamma(\kappa - 1)}{(4\pi)^{\kappa-1}}, \tag{10}$$

and where

$$\langle f, g \rangle_N = \int_{\Gamma_0(N) \backslash \mathbb{H}} y^\kappa f(z) \overline{g(z)} \frac{dx\,dy}{y^2}.$$

Since the main interest here is in the level aspect, we often suppress the dependence on the weight κ in the notation. The Petersson formula states

$$\Delta_N(m,n) = \delta_{m=n} + 2\pi i^{-\kappa} \sum_{c \equiv 0 (\mathrm{mod}\ N)} \frac{S(m,n;c)}{c} J_{\kappa-1}\left(\frac{4\pi\sqrt{mn}}{c}\right). \tag{11}$$

We have

Theorem 3 *Let N be square-free, and let $H_\kappa^*(N)$ denote the set of Hecke-normalized newforms on $\Gamma_0(N)$ of trivial central character. Let*

$$\Delta_N^*(m,n) = c_\kappa \sum_{f \in H_\kappa^*(N)} \frac{\overline{\lambda_f(m)}\lambda_f(n)}{\langle f, f\rangle_N}. \tag{12}$$

Then with $\nu(L)$ defined to be the completely multiplicative function satisfying $\nu(p) = p+1$ for p prime, we have

$$\Delta_N^*(m,n) = \sum_{LM=N} \frac{\mu(L)}{\nu(L)} \sum_{\ell | L^\infty} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 | \ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u|(m,L)\\ v|(n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \left| \left(\frac{m}{u}, \frac{u}{(u,v)}\right) e_1 \right| \left(d_1, \frac{m}{a^2(u,v)}\right)\\ b \left| \left(\frac{n}{v}, \frac{v}{(u,v)}\right) e_2 \right| \left(d_2, \frac{n}{b^2(u,v)}\right)}} \Delta_M\left(\frac{md_1}{a^2e_1^2(u,v)}, \frac{nd_2}{b^2e_2^2(u,v)}\right). \tag{13}$$

Here $c_\ell(d)$ with $d \mid \ell$ is jointly multiplicative, and $c_{p^n}(p^j) = c_{j,n}$ where

$$x^n = \sum_{j=0}^{n} c_{j,n} U_j\left(\frac{x}{2}\right), \tag{14}$$

and $U_j(x)$ are the Chebyshev polynomials of the second kind.

The constants $c_{j,n}$ arise from repeated application of the Hecke multiplicativity relations and we call them the *Chebyshev coefficients*. We describe some of their relevant properties in Sect. 6, for instance, we shall show $c_{j,n} \geq 0$, and derive sharp bounds on $c_{j,n}$. Many of the bounds on the Chebyshev coefficients appearing in Sect. 6 arose out of necessity for the proof of Theorem 1.

In Theorem 4, we give an approximate version of (13) with the additional restriction $\ell \leq Y$, which makes the right hand side a finite sum. For our application to the cubic moment, we have found the approximate version most suitable.

The method of proof of [26] is to explicitly choose a basis $B_k(N)$ (see [26, Proposition 2.6]) that relates the oldforms to the newforms, and thereby deduce an arithmetic formula for $\Delta_N(m,n)$ in terms of $\Delta_M^*(m',n')$'s, with $M \mid N$. An inversion of this formula then gives their formula for $\Delta_N^*(m,n)$. As mentioned in [26], there are many interesting choices of basis and it could be argued that their choice is ad-hoc. Other authors have also constructed various bases. Choie and Kohnen [12, Proposition 2] use

the same basis that we will use here. Rouymi [41] gave a basis for prime power level and derived a newform Petersson formula from it. Building on Rouymi, Ng [37] and Blomer and Milićević [11, (3.7)] gave a basis for arbitrary level, and Ng [37] and five authors [4] have used this basis to give newform Petersson formulas for arbitrary level (but with restrictive coprimality conditions on the level and the Fourier coefficients). It is important for our work that there is no restriction on m, n appearing in Theorem 3.

Nelson [36] has described a method for proving a Petersson formula for newforms without explicitly choosing a basis, and gives such a formula when the level N is divisible by the cube of each prime dividing it.

Our proof takes a different path from [4,11,26,37,41] in that we choose our basis to be eigenfunctions of the Atkin–Lehner operators, which for square-free level is enough to diagonalize the basis. This choice is natural and leads to many pleasant simplifications. Our method of proof of Theorem 3 most naturally shows

$$\Delta_N(m,n) = \sum_{LM=N} \frac{1}{\nu(L)} \sum_{\ell | L^\infty} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u|(m,L)\\ v|(n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \left| \left(\frac{m}{u}, \frac{u}{(u,v)}\right) e_1 \right| \left(d_1, \frac{m}{a^2(u,v)}\right) \\ b \left| \left(\frac{n}{v}, \frac{v}{(u,v)}\right) e_2 \right| \left(d_2, \frac{n}{b^2(u,v)}\right)}} \sum \Delta_M^* \left(\frac{md_1}{a^2 e_1^2 (u,v)}, \frac{nd_2}{b^2 e_2^2 (u,v)} \right). \tag{15}$$

We deduce Theorem 3 from (15) in Sect. 4 below.

In this introduction, we have not presented the Petersson formula that is required for the proof of Theorem 1. What we need is a kind of hybrid formula for modular forms of level rq that in the r-aspect restricts to newforms of level r, and in the q-aspect groups together all the newforms of level dividing q, in accordance with the setup of Theorem 1. This formula appears in Sect. 5.

The newform formula of [26, Proposition 2.8] has coprimality assumptions of the form $(m, N) = 1$ and $(n, N^2) \mid N$, which on the face of it is rather restrictive, however, one may reduce to this case as follows. Firstly, using that $\lambda_f(d)\lambda_f(p) = \lambda_f(dp)$ for any $d \in \mathbb{N}$, and $p \mid N$, one may write $\lambda_f(m)\lambda_f(n) = \lambda_f(m')\lambda_f(n')$ where $mn = m'n'$, and $(m', N) = 1$. Secondly, we have $\lambda_f(p^2 d) = \lambda_f(p)^2\lambda_f(d) = p^{-1}\lambda_f(d)$ (see (63)), which by repeated applications allows one to reach the case $(n, N^2) \mid N$. It is not obvious how to use the [26] newform formula to derive our hybrid version presented in Sect. 5 below. The problem is that the above factorizations of m and n depend on the ambient level, and so summing over different levels introduces some complications.

1.4 Structure of the paper

Sections 2–7 are devoted to proving a number of versions of the Petersson formula with newforms as well as some estimates for the Chebyshev coefficients. This part of the paper is self-contained.

In Sects. 8–11, we prove the cubic moment bound, that is, Theorem 1.

2 Atkin–Lehner theory

2.1 Construction of basis

We briefly review some of the theory developed by Atkin and Lehner [2]. Throughout we assume that the level N is square-free. For a matrix in $GL_2^+(\mathbb{Z})$, define

$$f|_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(z) = (ad-bc)^{\kappa/2}(cz+d)^{-\kappa} f\left(\frac{az+b}{cz+d}\right).$$

Atkin and Lehner showed that

$$S_\kappa(N) = \bigoplus_{LM=N} \bigoplus_{f \in H_\kappa^*(M)} S_\kappa(L; f),$$

where $S_\kappa(L; f)$ is the span of forms $f_{|A_\ell}$, with $\ell \mid L$, where $A_\ell = \begin{pmatrix} \ell & 0 \\ 0 & 1 \end{pmatrix}$. They call $S_\kappa(L; f)$ the *oldclass* associated to f. Observe $f_{|A_\ell}(z) = \ell^{\kappa/2} f(\ell z)$, so $S_\kappa(L; f) = \text{span}\{f(\ell z) : \ell \mid L\}$. Our goal here is to construct an explicit orthogonal basis of $S_\kappa(L; f)$, in the case that N is square-free.

We turn to the Atkin–Lehner operators W_d. Suppose that $d \mid N$, $N = LM$, and let

$$W_d = \begin{pmatrix} dx & y \\ Nz & dw \end{pmatrix}, \tag{16}$$

where $x, y, z, w \in \mathbb{Z}$ are chosen so that $\det(W_d) = d$ (such a choice exists because $(d, N/d) = 1$, since N is square-free, and the forthcoming properties of W_d are independent of the choices of x, y, z, w). If $d \mid M$ and $f \in H_\kappa^*(M)$ then f is an eigenfunction of W_d (see [2, Theorem 3 (iii)]), so suppose now that $d \mid L$. Let

$$V = \begin{pmatrix} 1 & 0 \\ \frac{-LM}{d} & 1 \end{pmatrix} \in \Gamma_0(M). \tag{17}$$

Note that taking $x = z = 1$ in the definition of W_d we have

$$VW_d = \begin{pmatrix} d & y \\ 0 & 1 \end{pmatrix}.$$

Therefore, if $f \in H_\kappa^*(M)$ and $d \mid L$, then

$$f_{|W_d} = f_{|V|W_d} = f_{|VW_d} = d^{\kappa/2} f(dz+y) = f_{|A_d}. \tag{18}$$

This calculation easily shows that $S_\kappa(L;f)$ is preserved by all W_d with $d \mid L$, that each W_d is an involution, and that the W_d commute with each other. Therefore, the group of transformations of $S_\kappa(L;f)$ generated by the W_d is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^{\omega(L)}$, where $\omega(n)$ is the number of prime divisors of n. Note $2^{\omega(L)} = \tau(L)$. Furthermore, the W_d are Hermitian with respect to the Petersson inner product (see [2, Lemma 25]). By some simple character theory, we can show that $S_\kappa(L;f)$ has an explicit orthogonal basis of common eigenfunctions of the W_d.

We briefly describe a more abstract statement. Let G be a group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^n$, and let ϕ be a character on G, which we denote by $\phi \in \widehat{G}$. There are 2^n such characters. For each $g \in G$, suppose there exists an involution W_g acting on some vector space of functions, and such that $W_{g_1}W_{g_2} = W_{g_1g_2}$. For each f in the vector space and character $\phi \in \widehat{G}$, define

$$f_\phi = \sum_{g\in G} \phi(g) W_g f. \tag{19}$$

It is easy to see that

$$W_g f_\phi = \phi(g) f_\phi.$$

Therefore, each f_ϕ is an eigenfunction of all the W_g. Also, the f_ϕ are distinct because any two choices of f_ϕ have a different eigenvalue for some W_g. This also means that if the W_g are Hermitian with respect to some inner product, then all the f_ϕ are orthogonal. In the case of $S_\kappa(L;f)$, which has dimension $2^{\omega(L)} = \tau(L)$, there are $2^{\omega(L)}$ eigenfunctions f_ϕ, so by dimension counting, the f_ϕ form a basis. Finally, we derive a useful formula for $\langle f_\phi, f_\phi\rangle$:

$$\begin{aligned}\langle f_\phi, f_\phi\rangle &= \sum_{g_1,g_2\in G} \phi(g_1)\phi(g_2)\langle W_{g_1}f, W_{g_2}f\rangle \\ &= \sum_{g_1,g_2\in G} \phi(g_1g_2)\langle W_{g_1g_2}f, f\rangle = |G|\sum_{g\in G}\phi(g)\langle W_g f, f\rangle. \end{aligned}\tag{20}$$

Returning to $S_\kappa(L;f)$, by [26, Lemma 2.4] (which in turn follows closely a proof in [1]), we have

$$\langle f_{|W_d}, f\rangle = \langle f_{|A_d}, f\rangle = \frac{\lambda_f(d)}{\nu(d)} d^{1/2}\langle f, f\rangle.$$

We endow the set of divisors $d \mid L$ with the group structure $(\mathbb{Z}/2\mathbb{Z})^{\omega(L)}$ and define characters on it by $\phi(d) = \prod_{p|d}\phi(p)$, where $\phi(p)$ is chosen to be $+1$ or -1 independently for each prime divisor of L. In this way, we obtain

$$\langle f_\phi, f_\phi\rangle = |G|\langle f, f\rangle \sum_{d|L}\phi(d)\frac{\lambda_f(d)}{\nu(d)}d^{1/2} = \tau(L)\langle f, f\rangle \prod_{p|L}\left(1 + \frac{\phi(p)\lambda_f(p)p^{1/2}}{\nu(p)}\right). \tag{21}$$

All of the above inner products are $\langle\,,\rangle_N$.

2.2 Dirichlet series of the basis of oldforms

To lend some support to the assertion that our choice of basis of $S_\kappa(L;\, f)$ given above is natural, here we describe some pleasant features of the Dirichlet series corresponding to these modular forms. Let $\lambda_{f_\phi}(n)$ be the Fourier coefficients of f_ϕ and define

$$D(s,\, f_\phi) = \sum_{n=1}^{\infty} \frac{\lambda_{f_\phi}(n)}{n^s}. \tag{22}$$

The reader should beware that this is not a character twist of f, because ϕ is not a Dirichlet character (in fact ϕ is only defined on the divisors of L). We show here that $D(s,\, f_\phi)$ satisfies a functional equation similar to that of a level N newform.

By a direct calculation with the Fourier expansion, we have

$$\lambda_{f_\phi}(m) = \sum_{u|(m,L)} \phi(u) u^{1/2} \lambda_f(m/u), \tag{23}$$

Therefore, we have

$$D(s,\, f_\phi) = L(s,\, f) \prod_{p|L} \left(1 + \frac{\phi(p)}{p^{s-1/2}}\right). \tag{24}$$

Then define the "completed" Dirichlet series

$$N^{s/2} \Gamma_f(s) D(s,\, f_\phi) = \Lambda(s,\, f) L^{s/2} \prod_{p|L} \left(1 + \frac{\phi(p)}{p^{s-1/2}}\right), \tag{25}$$

where

$$\Gamma_f(s) = \pi^{-s} \Gamma\left(\frac{s + (\kappa - 1)/2}{2}\right) \Gamma\left(\frac{s + (\kappa + 1)/2}{2}\right)$$

is the gamma factor associated to $L(s,\, f)$ and $\Lambda(s,\, f) = M^{s/2} \Gamma_f(s) L(s,\, f)$. This satisfies the functional equation $\Lambda(s,\, f) = \epsilon_f \Lambda(1 - s,\, f)$. Meanwhile, the secondary factor satisfies

$$g(s) := L^{s/2} \prod_{p|L} \left(1 + \frac{\phi(p)}{p^{s-1/2}}\right) = \prod_{p|L} \left(p^{s/2} + \phi(p) p^{\frac{1-s}{2}}\right) = \phi(L) g(1 - s).$$

Therefore, $D(s,\, f_\phi)$ satisfies the functional equation

$$N^{\frac{s}{2}} \Gamma_f(s) D(s,\, f_\phi) = \epsilon_f \phi(L) N^{\frac{1-s}{2}} \Gamma_f(1 - s) D(1 - s,\, f_\phi). \tag{26}$$

3 Manipulations with sums of Fourier coefficients

The goal of this section is to prove (15).

We begin by describing (10) for the basis chosen in Sect. 2.1. We have

$$\Delta_N(m,n) = c_\kappa \sum_{LM=N} \sum_{f\in H_\kappa^*(M)} \sum_{\phi} \frac{\overline{\lambda_{f_\phi}(m)}\lambda_{f_\phi}(n)}{\langle f_\phi, f_\phi\rangle_N}. \tag{27}$$

We therefore need to evaluate the inner sum over ϕ, namely

$$\begin{aligned} T(m,n) &:= \sum_{\phi} \frac{\overline{\lambda_{f_\phi}(m)}\lambda_{f_\phi}(n)}{\langle f_\phi, f_\phi\rangle_N} \\ &= \frac{1}{\tau(L)\langle f, f\rangle_N} \sum_{\phi} \overline{\lambda_{f_\phi}(m)}\lambda_{f_\phi}(n) \prod_{p|L} \left(1+\frac{\phi(p)\lambda_f(p)p^{1/2}}{\nu(p)}\right)^{-1}, \end{aligned} \tag{28}$$

where we have used (21). We multiply and divide by $\prod_{p|L}(1-\frac{\phi(p)\lambda_f(p)p^{1/2}}{\nu(p)})$, giving that

$$T(m,n) = \frac{1}{\tau(L)\rho_f(L)\langle f, f\rangle_N} \sum_{\phi} \overline{\lambda_{f_\phi}(m)}\lambda_{f_\phi}(n) \sum_{t|L} \frac{\mu(t)\phi(t)\lambda_f(t)t^{1/2}}{\nu(t)},$$

where as in [26], we define

$$\rho_f(L) = \prod_{p|L} \left(1 - p\frac{\lambda_f(p)^2}{(p+1)^2}\right). \tag{29}$$

The formula (23) implies

$$\begin{aligned} T(m,n) = \frac{1}{\tau(L)\rho_f(L)\langle f, f\rangle_N} &\sum_{\phi} \sum_{u|(m,L)} \sum_{v|(n,L)} \\ &\times \sum_{t|L} \phi(u)u^{1/2}\lambda_f(m/u)\phi(v)v^{1/2}\lambda_f(n/v)\frac{\mu(t)\phi(t)\lambda_f(t)t^{1/2}}{\nu(t)}, \end{aligned}$$

where we have used that $\lambda_f(n)$ is real to remove the complex conjugate symbols. The sum over ϕ detects if uvt is a square, precisely

$$\sum_{\phi} \phi(u)\phi(v)\phi(t) = \begin{cases} \tau(L), & \text{if } uvt = \square, \\ 0, & \text{otherwise.} \end{cases}$$

Since u and v are square-free, and so is t, the condition $uvt = \square$ determines t uniquely, namely

$$t = \frac{uv}{(u,v)^2} = \frac{u}{(u,v)}\frac{v}{(u,v)}.$$

One may compare this with Lemma 2.4 of [26]. Therefore, $T(m,n)$ equals

$$\frac{1}{\rho_f(L)\langle f,f\rangle_N}\sum_{u|(m,L)}\sum_{v|(n,L)} u^{1/2}\lambda_f(m/u)v^{1/2}\lambda_f(n/v)\frac{\mu(\frac{uv}{(u,v)^2})\lambda_f\left(\frac{uv}{(u,v)^2}\right)\left(\frac{uv}{(u,v)^2}\right)^{1/2}}{\nu\left(\frac{uv}{(u,v)^2}\right)}. \tag{30}$$

To check this against [26], suppose that $(m,n,N)=1$. This means $((m,L),(n,L)) = (m,n,L) \mid (m,n,N) = 1$, so in particular, $(u,v)=1$. Hence

$$T(m,n) = \frac{1}{\rho_f(L)\langle f,f\rangle_N}\sum_{u|(m,L)}\frac{\mu(u)u}{\nu(u)}\lambda_f(u)\lambda_f(m/u)\sum_{v|(n,L)}\frac{\mu(v)v}{\nu(v)}\lambda_f(v)\lambda_f(n/v), \tag{31}$$

which equals

$$\frac{A_f(m,L)A_f(n,L)}{\rho_f(L)\langle f,f\rangle_N}, \quad \text{where} \quad A_f(m,L) := \sum_{u|(m,L)}\frac{\mu(u)u}{\nu(u)}\lambda_f(u)\lambda_f(m/u),$$

as in [26, p.76]. Since f is on $\Gamma_0(M)$, we have $\langle f,f\rangle_N = \frac{\nu(N)}{\nu(M)}\langle f,f\rangle_M = \nu(L)\langle f,f\rangle_M$. So, if $(m,n,N)=1$, then

$$\Delta_N(m,n) = c_\kappa \sum_{LM=N}\sum_{f\in H_\kappa^*(M)}\frac{A_f(m,L)A_f(n,L)}{\rho_f(L)\nu(L)\langle f,f\rangle_M}. \tag{32}$$

From this we may quickly derive (2.48) of [26].

We continue with the calculation of $\Delta_N(m,n)$, without the assumption $(m,n,N)=1$. The formula (30) shows

$$\Delta_N(m,n) = c_\kappa \sum_{LM=N}\sum_{f\in H_\kappa^*(M)}\frac{1}{\nu(L)\rho_f(L)\langle f,f\rangle_M} \times \sum_{\substack{u|(m,L)\\ v|(n,L)}}\frac{uv}{(u,v)}\frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}\lambda_f\left(\frac{m}{u}\right)\lambda_f\left(\frac{u}{(u,v)}\right)\lambda_f\left(\frac{n}{v}\right)\lambda_f\left(\frac{v}{(u,v)}\right).$$

Recall that the Hecke relation for a newform of level M with trivial nebentypus is

$$\lambda_f(m)\lambda_f(n) = \sum_{\substack{d|(m,n)\\ (d,M)=1}}\lambda_f\left(\frac{mn}{d^2}\right).$$

In our desired application, u and v divide L and $(M, L) = 1$, so any divisor of u or v is automatically coprime to M. Using the Hecke relation, we then deduce

$$\Delta_N(m,n) = c_\kappa \sum_{LM=N} \sum_{f \in H_\kappa^*(M)} \frac{1}{\nu(L)\rho_f(L)\langle f, f\rangle_M} \times \sum_{\substack{u|(m,L)\\ v|(n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)} \sum_{\substack{a \left| \left(\frac{m}{u}, \frac{u}{(u,v)}\right)\\ b \left| \left(\frac{n}{v}, \frac{v}{(u,v)}\right)\right.}} \lambda_f\left(\frac{m}{a^2(u,v)}\right) \lambda_f\left(\frac{n}{b^2(u,v)}\right). \tag{33}$$

The tricky part of our analysis of $\Delta_N(m,n)$ is to express $\rho_f(L)^{-1}$ in terms of Fourier coefficients of f. We have

$$\frac{1}{\rho_f(L)} = \prod_{p|L} \left(1 - p\frac{\lambda_f(p)^2}{(p+1)^2}\right)^{-1} = \sum_{\ell | L^\infty} \ell \frac{\lambda_f^*(\ell)^2}{\nu(\ell)^2}, \tag{34}$$

where $\lambda_f^*(\ell)$ is the completely multiplicative version of $\lambda_f(n)$, that is,

$$\lambda_f^*(\ell) = \prod_{p^n || \ell} \lambda_f(p)^n.$$

Using only the weak bound $|\lambda_f(p)| \le p^\theta + p^{-\theta}$ with some $\theta < 1/2$ shows that the product and sum in (34) converge absolutely.

Define the *Chebyshev coefficients* $c_{j,n}$ by

$$\lambda_f(p)^n = \sum_{j=0}^{n} c_{j,n} \lambda_f(p^j), \tag{35}$$

where p is coprime to the level of f. Let $U_k(x)$ denote the degree k Chebyshev polynomial of the second kind (defined below). Then

$$x^n = \sum_{j=0}^{n} c_{j,n} U_j\left(\frac{x}{2}\right),$$

where the $c_{j,n}$ can be written in various ways using that the U_j form a system of orthogonal polynomials. Here the U_j can be defined concisely by the generating function

$$(1 - 2yx + x^2)^{-1} = \sum_{j=0}^{\infty} U_j(y) x^j.$$

For instance, since $U_4(x/2) = x^4 - 3x^2 + 1$, $U_2(x/2) = x^2 - 1$, and $U_0(x) = 1$, we get

$$x^4 = U_4(x/2) + 3U_2(x/2) + 2U_0(x/2).$$

An alternative formula is $U_j(\cos(\theta)) = \frac{\sin((j+1)\theta)}{\sin\theta}$. The orthogonality of the U_j implies that

$$c_{j,n} = \int_0^\pi U_j(\cos\theta)(2\cos\theta)^n \tfrac{2}{\pi}\sin^2\theta d\theta. \tag{36}$$

We will develop some properties of the Chebyshev coefficients in Sect. 6.

With this notation in hand, we have for f a newform of level M with $(\ell, M) = 1$, that

$$\lambda_f^*(\ell) = \prod_{p^n||\ell}\left(\sum_{j=0}^n c_{j,n}\lambda_f(p^j)\right) =: \sum_{d|\ell} c_\ell(d)\lambda_f(d), \tag{37}$$

where

$$c_\ell(d) = \prod_{\substack{p^j||d\\ p^n||\ell}} c_{j,n}.$$

Moreover, we have

$$\frac{1}{\rho_f(L)} = \sum_{\ell|L^\infty}\frac{\ell}{\nu(\ell)^2}\sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2)\lambda_f(d_1)\lambda_f(d_2). \tag{38}$$

Inserting (38) into (33), we get

$$\begin{aligned}\Delta_N(m,n) = c_\kappa \sum_{LM=N}\sum_{\ell|L^\infty}\frac{\ell}{\nu(\ell)^2\nu(L)}\sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2)\sum_{f\in H_\kappa^*(M)}\frac{1}{\langle f,f\rangle_M}\\ \times\lambda_f(d_1)\lambda_f(d_2)\sum_{\substack{u|(m,L)\\ v|(n,L)}}\frac{uv}{(u,v)}\frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}\sum_{\substack{a\left|\left(\frac{m}{u},\frac{u}{(u,v)}\right)\\ b\left|\left(\frac{n}{v},\frac{v}{(u,v)}\right)\right.\right.}}\lambda_f\left(\frac{m}{a^2(u,v)}\right)\lambda_f\left(\frac{n}{b^2(u,v)}\right).\end{aligned}$$

Now we can use the Hecke relations one final time (again the divisors are coprime to M), to give

$$\begin{aligned}\Delta_N(m,n) = c_\kappa \sum_{LM=N}\sum_{\ell|L^\infty}\frac{\ell}{\nu(\ell)^2\nu(L)}\sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2)\sum_{f\in H_\kappa^*(M)}\frac{1}{\langle f,f\rangle_M}\\ \times\sum_{\substack{u|(m,L)\\ v|(n,L)}}\frac{uv}{(u,v)}\frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}\sum_{\substack{a\left|\left(\frac{m}{u},\frac{u}{(u,v)}\right)\\ b\left|\left(\frac{n}{v},\frac{v}{(u,v)}\right)\right.\right.}}\sum_{\substack{e_1\left|\left(d_1,\frac{m}{a^2(u,v)}\right)\\ e_2\left|\left(d_2,\frac{n}{b^2(u,v)}\right)\right.\right.}}\lambda_f\left(\frac{md_1}{a^2e_1^2(u,v)}\right)\lambda_f\left(\frac{nd_2}{b^2e_2^2(u,v)}\right).\end{aligned}$$

This is precisely the desired formula (15), after a rearrangement.

4 Inversion

In this section, we show how to deduce (13) from (15). We work in greater generality than what is immediately required, which will be useful in Sect. 5. Suppose that F and G are two arithmetic functions that we write in the form $F(m, N)$, $G(m, N)$ where N is a positive square-free integer, and $m = (m_1, \dots, m_d)$ is a tuple of integers. We assume there is a relation of the form

$$F(m, N) = \sum_{LM=N} \sum_{a | L^\infty} A(a, m_L) G\left(\alpha(a, m_L) \frac{m}{m_L}, M\right), \tag{39}$$

where: A is some multiplicative arithmetical function, m_L denotes the part of m having common factors with L, so $m_L \mid L^\infty$ and $(m/m_L, L) = 1$, and α is some integer-valued multiplicative function having the property that $\alpha(p^{i_1}, \dots, p^{i_k}) \mid p^\infty$ for all primes p. Furthermore, a is shorthand for some tuple $(a_1, \dots, a_J)$, and the condition $a \mid L^\infty$ means that $a_i \mid L^\infty$ for all $i = 1, \dots, J$.

We can derive that $F(m, 1) = G(m, 1)$ for all m, by taking $N = 1$ in (39). The main topic of this section is to prove

Lemma 1 *For square-free N we have the inversion formula*

$$G(m, N) = \sum_{LM=N} \mu(L) \sum_{a | L^\infty} A(a, m_L) F\left(\alpha(a, m_L) \frac{m}{m_L}, M\right). \tag{40}$$

Lemma 1 implies Theorem 3, since (39) encompasses (15), as we now explain. The tuple $(m_1, \dots m_d)$ in (39) takes the form (m, n) in (15). The tuple a appearing in (39) is of the form $(\ell, d_1, d_2, u, v, a, b, e_1, e_2)$, where note that all these entries divide L^∞. The arithmetical function $A(a, m_L)$ accounts for $\frac{1}{\nu(L)}, \frac{\ell}{\nu(\ell)^2}, \dots, \frac{1}{\nu(uv/(u,v)^2)}$, as well as all the summation conditions in (15), in which (m, n) may be replaced by (m_L, n_L). Finally, we have $\alpha(a, (m_L, n_L)) = (\frac{m_L d_1}{a^2 e_1^2 (u,v)}, \frac{n_L d_2}{b^2 e_2^2 (u,v)})$, and $G(x, y, M) = \Delta_M^*(x, y)$.

Proof If $N = 1$, then (40) is true, by an easy calculation.

Now induct on the number of prime factors of N. We replace N by NP with P a prime (whence $(P, N) = 1$), giving

$$F(m, NP) = G(m, NP) + \sum_{\substack{LM=NP \\ L \neq 1}} \sum_{a | L^\infty} A(a, m_L) G\left(\alpha(a, m_L) \frac{m}{m_L}, M\right).$$

Since M has fewer prime factors than NP, we can use the induction hypothesis to give

$$F(m,NP) = G(m,NP) + \Bigg[\sum_{\substack{LM=NP \\ L\neq 1}} \sum_{a|L^\infty} A(a,m_L) \times \sum_{CD=M} \mu(C) \sum_{b|C^\infty} A(b,m_C) F\left(\alpha(b,m_C)\frac{\alpha(a,m_L)\frac{m}{m_L}}{m_C}, D\right)\Bigg].$$

Here we have used that

$$\left(\alpha(a,m_L)\frac{m}{m_L}\right)_C = m_C,$$

which follows from $(L,C)=1$. Next we put back $L=1$ and subtract it off again, giving

$$F(m,NP) = G(m,NP) - \sum_{CD=NP} \mu(C) \sum_{b|C^\infty} A(b,m_C) F\left(\alpha(b,m_C)\frac{m}{m_C}, D\right) + \Bigg[\sum_{LM=NP} \sum_{a|L^\infty} A(a,m_L) \sum_{CD=M} \mu(C) \sum_{b|C^\infty} A(b,m_C) F\left(\alpha(b,m_C)\frac{\alpha(a,m_L)\frac{m}{m_L}}{m_C}, D\right)\Bigg].$$

We need to show that the term in square brackets equals $F(m,NP)$, since we can then solve for $G(m,NP)$, giving (40). We have

$$[\dots] = \sum_{CDL=NP} \mu(C) \sum_{a|L^\infty} A(a,m_L) \sum_{b|C^\infty} A(b,m_C) F\left(\alpha(b,m_C)\frac{\alpha(a,m_L)\frac{m}{m_L}}{m_C}, D\right).$$

Using multiplicativity of A and α, and that $(C,L)=1$, we get

$$[\dots] = \sum_{CDL=NP} \mu(C) \sum_{c|(LC)^\infty} A(c,m_{CL}) F\left(\alpha(c,m_{CL})\frac{m}{m_{CL}}, D\right).$$

We can write this as

$$[\dots] = \sum_{D|NP} \sum_{c|(NP/D)^\infty} A(c,m_{NP/D}) F(\alpha(c,m_{NP/D})\frac{m}{m_{NP/D}}, D) \sum_{CL=NP/D} \mu(C).$$

The inner sum over C gives $D=NP$, which simplifies as $[\dots] = F(m,NP)$, as desired. □

5 Hybrid formulas

We desire a formula that is intermediate between $\Delta_N(m,n)$ and $\Delta_N^*(m,n)$, in order to capture the weights appearing in Theorem 1. See the discussion surrounding (4) for motivation for this goal.

For square-free AB, define

$$\widetilde{\Delta}_{A,B}(m,n) = c_\kappa \sum_{CD=B} \sum_{f \in H_\kappa^*(AD)} \frac{1}{\nu(C)} \frac{1}{\rho_f(C)} \frac{\lambda_f(m)\lambda_f(n)}{\langle f, f\rangle_{AD}}. \tag{41}$$

Note that $\widetilde{\Delta}_{A,1}(m,n) = \Delta_A^*(m,n)$. Provided $(mn, B) = 1$, we have $A_f(m, C) = \lambda_f(m)$ for m coprime to C, and then by (32) and we get $\widetilde{\Delta}_{1,B}(m,n) = \Delta_B(m,n)$. Hence $\widetilde{\Delta}_{A,B}(m,n)$ interpolates between $\Delta_{AB}(m,n)$ and $\Delta_{AB}^*(m,n)$, provided $(mn, B) = 1$.

Because of our application to the cubic moment problem, we are interested in the case where qN is square-free, and $(mn, q) = 1$. In (33) we substitute $N \to Nq$, and factor $L = L_N L_q$ with $L_N \mid N$ and $L_q \mid q$, giving

$$\Delta_{Nq}(m,n) = c_\kappa \sum_{L_N \mid N} \frac{1}{\nu(L_N)} \sum_{\substack{u \mid (m, L_N) \\ v \mid (n, L_N)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)} \sum_{\substack{a \mid \left(\frac{m}{u}, \frac{u}{(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right)}}$$
$$\times \sum_{\substack{L_N L_q M = Nq \\ L_q \mid q}} \sum_{f \in H_\kappa^*(M)} \frac{\lambda_f\left(\frac{m}{a^2(u,v)}\right) \lambda_f\left(\frac{n}{b^2(u,v)}\right)}{\nu(L_q)\rho_f(L_q)\rho_f(L_N)\langle f, f\rangle_M}.$$

Here we used that $(mn, L_q) = 1$ to simplify the divisiblity conditions. Next we use (38) on $\frac{1}{\rho_f(L_N)}$, and use the Hecke relation again, giving now

$$\Delta_{Nq}(m,n) = \sum_{L_N \mid N} \sum_{\ell \mid L_N^\infty} \frac{\ell}{\nu(L_N)\nu(\ell)^2} \sum_{d_1, d_2 \mid \ell} c_\ell(d_1) c_\ell(d_2) \sum_{\substack{u \mid (m, L_N) \\ v \mid (n, L_N)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \mid \left(\frac{m}{u}, \frac{u}{(u,v)}\right) e_1 \mid \left(d_1, \frac{m}{a^2(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right) e_2 \mid \left(d_2, \frac{n}{b^2(u,v)}\right)}} c_\kappa \sum_{L_q \mid q} \sum_{f \in H_\kappa^*\left(\frac{N}{L_N}, \frac{q}{L_q}\right)} \frac{\lambda_f\left(\frac{md_1}{a^2 e_1^2(u,v)}\right) \lambda_f\left(\frac{nd_2}{b^2 e_2^2(u,v)}\right)}{\nu(L_q)\rho_f(L_q)\langle f, f\rangle_{\frac{Nq}{L_N L_q}}}.$$

With $A = N/L_N$, $B = q$, $C = L_q$, and $D = q/L_q$, we can write the sum over $L_q \mid q$ above as $\widetilde{\Delta}_{N/L_N, q}(m', n')$, for m' and n' the obvious integers. Therefore, this shows

$$\Delta_{Nq}(m,n) = \sum_{L_N \mid N} \sum_{\ell \mid L_N^\infty} \frac{\ell}{\nu(L_N)\nu(\ell)^2} \sum_{d_1, d_2 \mid \ell} c_\ell(d_1) c_\ell(d_2) \sum_{\substack{u \mid (m, L_N) \\ v \mid (n, L_N)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \mid \left(\frac{m}{v}, \frac{u}{(u,v)}\right) e_1 \mid \left(d_1, \frac{m}{a^2(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right) e_2 \mid \left(d_2, \frac{n}{b^2(u,v)}\right)}} \widetilde{\Delta}_{\frac{N}{L_N}, q}\left(\frac{md_1}{a^2 e_1^2(u,v)}, \frac{nd_2}{b^2 e_2^2(u,v)}\right).$$

At this point, we replace the condition $L_N \mid N$ by $LM = N$ (i.e. rename L_N by just L, and then call M the complementary divisor). This gives

$$\Delta_{Nq}(m,n) = \sum_{LM=N} \sum_{\ell \mid L^\infty} \frac{\ell}{\nu(L)\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u \mid (m,L) \\ v \mid (n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \mid \left(\frac{m}{u}, \frac{u}{(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right)}} \sum_{\substack{e_1 \mid \left(d_1, \frac{m}{a^2(u,v)}\right) \\ e_2 \mid \left(d_2, \frac{n}{b^2(u,v)}\right)}} \widetilde{\Delta}_{M,q}\left(\frac{md_1}{a^2e_1^2(u,v)}, \frac{nd_2}{b^2e_2^2(u,v)}\right). \tag{42}$$

Now we fix q, and suppress it in the following notation. Set $F(m,n,N) = \Delta_{Nq}(m,n)$, and likewise $G(m,n,N) = \widetilde{\Delta}_{N,q}(m,n)$. Then the relation (42) is essentially the same formula as (15), and, more precisely, is encompassed by (39). By Lemma 1, we therefore have (recall $(mn,q) = 1$, and Nq is square-free)

$$\widetilde{\Delta}_{N,q}(m,n) = \sum_{LM=N} \frac{\mu(L)}{\nu(L)} \sum_{\ell \mid L^\infty} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u \mid (m,L) \\ v \mid (n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \mid \left(\frac{m}{u}, \frac{u}{(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right)}} \sum_{\substack{e_1 \mid \left(d_1, \frac{m}{a^2(u,v)}\right) \\ e_2 \mid \left(d_2, \frac{n}{b^2(u,v)}\right)}} \Delta_{Mq}\left(\frac{md_1}{a^2e_1^2(u,v)}, \frac{nd_2}{b^2e_2^2(u,v)}\right). \tag{43}$$

6 Formulas and estimates with Chebyshev coefficients

We begin with a combinatorial evaluation of $c_{j,n}$. From (36), combined with the formula $U_j(\cos\theta) = \frac{\sin((j+1)\theta)}{\sin\theta}$, we have

$$c_{j,n} = \tfrac{1}{2} \int_{-\pi}^{\pi} \sin((j+1)\theta)(2\cos(\theta))^n \tfrac{2}{\pi} \sin\theta d\theta. \tag{44}$$

Writing everything in terms of $e^{i\theta}$, we get

$$c_{j,n} = \frac{1}{\pi} \frac{1}{(2i)^2} \int_{-\pi}^{\pi} \left(\sum_{r=0}^{n} \binom{n}{r} e^{i\theta(n-2r)} (e^{i(j+2)\theta} - e^{ij\theta} - e^{-ij\theta} + e^{-i(j+2)\theta} \right) d\theta.$$

At this point it is clear that $c_{j,n} = 0$ if $j \not\equiv n \pmod 2$, so assume $j \equiv n \pmod 2$. We also see that $c_{j,n} = 0$ if $j > n$. For $j = n$, it is easy to check that $c_{n,n} = 1$, since the only values of r for which the integral does not vanish are $r = 0$, and $r = n$. If $j \leq n-2$, then we deduce that

$$c_{j,n} = \binom{n}{\frac{n+j}{2}} - \binom{n}{\frac{n+j}{2}+1}, \tag{45}$$

which in fact agrees with $c_{n,n}$ too, since $\binom{n}{n+1} = 0$. One may find this sequence of Chebyshev coefficients in the OEIS [38] which thereby leads to interesting connections. For instance, the list (where we omit terms with $j \not\equiv n \pmod 2$)

$$c_{0,0}, c_{1,1}, c_{0,2}, c_{2,2}, c_{1,3}, c_{3,3}, c_{0,4}, c_{2,4}, c_{4,4}, \ldots = 1, 1, 1, 1, 2, 1, 2, 3, 1, 5, 4, 1, 5 \ldots$$

is the same as Catalan's triangle ordered along diagonals in reverse order.

From (45), we deduce $c_{j,n} \geq 0$ for all j, n, and we have

$$\sum_{j=0}^{n} c_{j,n} = \begin{cases} \binom{n}{n/2}, & n \text{ even}, \\ \binom{n}{(n+1)/2}, & n \text{ odd}, \end{cases} \tag{46}$$

since the sum telescopes. Let $\delta \in \{0, 1\}$, $\delta \equiv n \pmod 2$. Note that Stirling's formula gives

$$\binom{n}{\frac{n+\delta}{2}} \sim \frac{2^n}{\sqrt{\pi n/2}}. \tag{47}$$

Lemma 2 *Let $c_\ell(d)$ be as defined in Sect. 3, and suppose $\gamma \geq 0$. Then*

$$\sum_{d|\ell} c_\ell(d) d^\gamma \leq \prod_{p^n \| \ell} (p^{-\gamma} + p^\gamma)^n.$$

Remark For $\gamma = 0$, this bound is slightly worse than that implied by (46), in view of (47).

Proof We have

$$\sum_{d|\ell} c_\ell(d) d^\gamma = \prod_{p^n \| \ell} \sum_{j=0}^{n} p^{\gamma j} c_{j,n}, \tag{48}$$

so it suffices to show for $x > 0$ that

$$\sum_{j=0}^{n} c_{j,n} x^j \leq (x^{-1} + x)^n. \tag{49}$$

From (45), we obviously have that

$$c_{j,n} \leq \binom{n}{\frac{n+j}{2}}. \tag{50}$$

From this, we shall deduce (49), as we presently explain. On the one hand, we have from (50) that

$$\sum_{j=0}^{n} c_{j,n} x^j \leq x^n + \binom{n}{n-1} x^{n-2} + \cdots + \binom{n}{\frac{n+\delta}{2}} x^{\delta}, \tag{51}$$

and on the other hand, we have

$$(x^{-1}+x)^n = x^n + \binom{n}{n-1} x^{n-2} + \cdots + \binom{n}{\frac{n+\delta}{2}} x^{\delta} + \binom{n}{\frac{n+\delta}{2}-1} x^{\delta-2} + \cdots + x^{-n}. \tag{52}$$

In words, the positive powers of x appearing in (52) precisely match the upper bound on (51). □

For later use, we require an estimate for the following expression:

$$S(L,Y) := \sum_{\substack{\ell | L^\infty \\ \ell \leq Y}} \frac{\ell}{\nu(\ell)^2} \left(\sum_{d|\ell} c_\ell(d) d^{1/2} \right)^2. \tag{53}$$

Lemma 3 *We have*

$$S(L,Y) \ll_\varepsilon Y^\varepsilon \tau(L). \tag{54}$$

Proof Without the restriction $\ell \leq Y$, the estimate in Lemma 2 would barely fail to show the sum converges, since

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

However, using Rankin's trick and Lemma 2 we obtain

$$S(L,Y) \leq \sum_{\ell | L^\infty} \left(\frac{Y}{\ell}\right)^\varepsilon = Y^\varepsilon \prod_{p|L} (1-p^{-\varepsilon})^{-1} \leq C(\varepsilon) Y^\varepsilon \tau(L),$$

where we have used the following:

$$\prod_{p|L} (1-p^{-\varepsilon})^{-1} \leq \left(\prod_{\substack{p|L \\ p^\varepsilon \geq 2}} 2 \right) \left(\prod_{\substack{p|L \\ p^\varepsilon < 2}} (1-p^{-\varepsilon})^{-1} \right) \leq \tau(L) C(\varepsilon), \tag{55}$$

where $C(\varepsilon) = \prod_{p: p^\varepsilon < 2} (1-p^{-\varepsilon})^{-1}$. □

We will additionally need the following bound.

Lemma 4 *Let $c_\ell(d)$ be as defined in Sect. 3. Then*

$$\frac{\sum_{d|\ell} c_\ell(d)}{\nu(\ell)} \leq \ell^{-\frac{\log(3/2)}{\log 2}} = \ell^{-0.5849\ldots}. \tag{56}$$

Proof By Lemma 2, we have

$$\ell^{\frac{\log(3/2)}{\log 2}} \frac{\sum_{d|\ell} c_\ell(d)}{\nu(\ell)} \leq \prod_{p^n||\ell} \left(\frac{2}{(p+1)} p^{\frac{\log(3/2)}{\log 2}} \right)^n. \tag{57}$$

Let

$$f_\delta(x) = \frac{2x^\delta}{x+1},$$

where $x \geq 2$ and $0 < \delta < 1$. If $\delta = \frac{\log(3/2)}{\log 2}$ is such that $f_\delta(x)$ is decreasing for $x \geq 2$, and $f_\delta(2) \leq 1$, then this will show that the product on the right hand side of (57) is ≤ 1, which suffices to prove the desired bound. It is easy to check with basic calculus that the desired properties hold for $f_\delta(x)$. □

Remark The exponent occurring in (56) is mainly controlled by the powers of 2 dividing ℓ. If $\ell = 2^n$, and n is even, then in fact $\sum_{d|\ell} c_\ell(d) = \binom{n}{n/2} \asymp 2^n/n^{1/2}$, while $\nu(\ell) = 3^n$, so if $\ell = 2^n$ then (56) is sharp up to the factor $n^{-1/2} = (\log_2 \ell)^{-1/2}$. If ℓ has no small prime divisors then the exponent can be improved.

Corollary 2 *Let $\gamma_0 = \frac{\log(3/2)}{\log 2} - \frac{1}{2} = 0.0849625\ldots$, and suppose $\varepsilon > 0$ is small. Then*

$$\sum_{\substack{\ell | L^\infty \\ \ell > Y}} \frac{\ell^{1+\varepsilon}}{\nu(\ell)^2} \left(\sum_{d|\ell} c_\ell(d) \right)^2 \ll_\varepsilon Y^{-2\gamma_0+2\varepsilon} \tau(L). \tag{58}$$

Proof By Lemma 4, we have

$$\frac{\ell^{1+\varepsilon}}{\nu(\ell)^2} \left(\sum_{d|\ell} c_\ell(d) \right)^2 \leq \ell^{\varepsilon - 2\gamma_0}.$$

Then, we have by Rankin's trick and (55), that

$$\sum_{\substack{\ell | L^\infty \\ \ell > Y}} \ell^{-2\gamma_0+\varepsilon} \leq Y^{-2\gamma_0+2\varepsilon} \sum_{\ell | L^\infty} \ell^{-\varepsilon} = Y^{-2\gamma_0+2\varepsilon} \prod_{p|L} (1-p^{-\varepsilon})^{-1} \ll_\varepsilon Y^{-2\gamma_0+2\varepsilon} \tau(L).$$

□

7 Approximate Petersson formulas

Our main Petersson formula, (13), has a technical problem arising from the fact that the sum over $\ell \mid L^\infty$ is not a finite sum. Iwaniec, Luo, and Sarnak encountered a similar difficulty in [26]. The idea is to truncate the sum at some large parameter Y, and estimate the tail trivially.

To this end, we begin with some simple bounds. Throughout this section we assume the weight κ is fixed, and do not display any κ-dependence in the error terms. First, we claim the crude bound

$$|\Delta_N(m,n)| \ll_\kappa (m,N)^{1/2}(n,N)^{1/2}\tau_3(m)\tau_3(n), \tag{59}$$

for fixed weight κ.

Proof We use (27). Using (23) and the Deligne bound, we have

$$|\lambda_{f_\phi}(m)| \le \sum_{u|(m,L)} u^{1/2}\tau(m/u) \le (m,L)^{1/2}\tau_3(m) \le (m,N)^{1/2}\tau_3(m).$$

Therefore, by the fact that $\lambda_{f_\phi}(1) = \lambda_f(1) = 1$, we have

$$|\Delta_N(m,n)| \le (m,N)^{1/2}(n,N)^{1/2}\tau_3(m)\tau_3(n)\Delta_N(1,1).$$

One can then apply the Kloosterman sum formula for $\Delta_N(1,1)$ to show (e.g. see [26, Corollary 2.3])

$$\Delta_N(1,1) = 1 + O\left(\frac{\tau(N)}{N^{3/2}}\right) \ll 1.$$

□

Now we state an approximate version of Theorem 3.

Theorem 4 *Let* $\gamma_0 = \frac{\log(3/2)}{\log 2} - \frac{1}{2} = 0.0849625\ldots$, *and suppose* $\varepsilon > 0$ *is small. We have*

$$\begin{aligned}\Delta_N^*(m,n) = {} & \sum_{LM=N} \frac{\mu(L)}{\nu(L)} \sum_{\substack{\ell \mid L^\infty \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u \mid (m,L) \\ v \mid (n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)} \\ & \times \sum_{\substack{a \mid \left(\frac{m}{u}, \frac{u}{(u,v)}\right) e_1 \mid \left(d_1, \frac{m}{a^2(u,v)}\right) \\ b \mid \left(\frac{n}{v}, \frac{v}{(u,v)}\right) e_2 \mid \left(d_2, \frac{n}{b^2(u,v)}\right)}} \Delta_M\left(\frac{md_1}{a^2e_1^2(u,v)}, \frac{nd_2}{b^2e_2^2(u,v)}\right) \\ & + O((mnNY)^\varepsilon NY^{-2\gamma_0}).\end{aligned} \tag{60}$$

Proof It suffices to bound the tail of the sum over ℓ, namely the terms with $\ell > Y$. Using (59), we have that the difference between $\Delta_N^*(m,n)$ and the main term sum on the right hand side of (60) is

$$\ll \sum_{LM=N} \frac{1}{\nu(L)} \sum_{\substack{\ell | L^\infty \\ \ell > Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u|(m,L) \\ v|(n,L)}} \frac{uv}{(u,v)} \frac{1}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \left| \left(\frac{m}{u}, \frac{u}{(u,v)}\right) \\ b \left| \left(\frac{n}{v}, \frac{v}{(u,v)}\right)}} \sum_{\substack{e_1 \left| \left(d_1, \frac{m}{a^2(u,v)}\right) \\ e_2 \left| \left(d_2, \frac{n}{b^2(u,v)}\right)}} \left(M, \frac{md_1}{a^2 e_1^2 (u,v)}\right)^{1/2} \left(M, \frac{nd_2}{b^2 e_2^2 (u,v)}\right)^{1/2} \tau_3(md_1)\tau_3(nd_2). \tag{61}$$

We use the weak bound $(M, m') \le M$ (for any integer m'), and use $\tau_3(md_1) \ll (m\ell)^\varepsilon$, and similarly for $\tau_3(nd_2)$, and trivially estimate the sums over u, v, a, b, e_1, e_2 to give that (61) is

$$\ll (mn)^\varepsilon N \sum_{L|N} \frac{L^\varepsilon}{\nu(L)} \sum_{\substack{\ell | L^\infty \\ \ell > Y}} \frac{\ell^{1+\varepsilon}}{\nu(\ell)^2} \left(\sum_{d|\ell} c_\ell(d) \right)^2.$$

The desired bound then follows from Corollary 2. □

The same method of proof applies verbatim to $\widetilde{\Delta}_{N,q}(m,n)$:

Theorem 5 *Suppose* $(mnN, q) = 1$. *We have*

$$\widetilde{\Delta}_{N,q}(m,n) = \sum_{LM=N} \frac{\mu(L)}{\nu(L)} \sum_{\substack{\ell | L^\infty \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2|\ell} c_\ell(d_1)c_\ell(d_2) \sum_{\substack{u|(m,L) \\ v|(n,L)}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}$$
$$\times \sum_{\substack{a \left| \left(\frac{m}{u}, \frac{u}{(u,v)}\right) \\ b \left| \left(\frac{n}{v}, \frac{v}{(u,v)}\right)}} \sum_{\substack{e_1 \left| \left(d_1, \frac{m}{a^2(u,v)}\right) \\ e_2 \left| \left(d_2, \frac{n}{b^2(u,v)}\right)}} \Delta_{Mq}\left(\frac{md_1}{a^2 e_1^2 (u,v)}, \frac{nd_2}{b^2 e_2^2 (u,v)}\right)$$
$$+ O((mnNY)^\varepsilon N Y^{-2\gamma_0}). \tag{62}$$

In our desired application, we shall take Y to be a very large power of the level, in which case the error term is very small. For this reason, we made no attempt to optimize the error term.

8 Initial structural steps

8.1 Invariants of the twisted *L*-functions

We begin by calculating the root number and conductor of $L(s, f \otimes \chi_q)$, which is apparently somewhat difficult to locate in the literature. Our proof of Theorem 1 does not require any formula for the root number of the twisted L-function, but it is helpful for interpreting Corollary 1.

More generally, suppose that g is a weight κ newform of level N with trivial central character. Also recall the definition of the Atkin–Lehner operators (16). Then g is an eigenform for the W_d, and we write its eigenvalue as $g|_{W_d} = \eta_d(g)g$. Then the sign of the functional equation for $\Lambda(s, g)$ is given by $i^\kappa \eta_N(g)$. Since the Atkin–Lehner operators satisfy $g|_{W_{d_1}}|_{W_{d_2}} = g|_{W_{d_1 d_2}}$ for $(d_1, d_2) = 1$, it suffices to consider the eigenvalues of Atkin–Lehner operartors $\eta_Q(g)$ where Q is a prime power dividing the conductor of g.

Let χ_q be a primitive quadratic Dirichlet character of conductor $q = q_o q_e$ with q_o odd and q_e a power of 2. Explicitly, $\chi_q(n) = \left(\frac{n}{q_o}\right)\chi_{q_e}(n)$ where $\left(\frac{n}{q_o}\right)$ is the Jacobi symbol and $\chi_{q_e}(n)$ is either 1, χ_4, or one of the two primitive quadratic characters of conductor 8. Recall we set $\tilde{q} = \mathrm{rad}(q)$ the largest square-free divisor of q. Let f be a newform of square-free level rq', where $(r, q) = 1$ and $q' \mid \tilde{q}$. We also take $q'' \mid q$ to be such that $q'' \mid q'^{\infty}$ and $(q/q'', q') = 1$. Let us write $f \otimes \chi_q = (f \otimes \chi_{q''}) \otimes \chi_{q/q''}$. We have by [3, Theorem 4.1] and e.g. [25, Proposition 14.20] that $f \otimes \chi_{q''}$ and $f \otimes \chi_q$ are newforms of conductors rq''^2 and rq^2, respectively.

We have by [15, (5.5.1)] that for each $p \mid r$ that

$$\eta_p(f \otimes \chi_{q''}) = \chi_{q''}(p)\eta_p(f)$$

where in Deligne's notation $a(V) = 1$ by our square-free hypothesis on r and $\dim(V) = 2$. We found the exposition by Pacetti [39] particularly helpful for these calculations. For each $p \mid q''$ we write P for the power of p dividing q''. Now we have

$$\eta_{P^2}(f \otimes \chi_{q''}) = \chi_P(-1),$$

by Atkin-Li Theorem 4.1 [3], and where we have written $\chi_{q''} = \prod_{p|q'} \chi_P$. Therefore we have shown that

$$\eta_{rq''^2}(f \otimes \chi_{q''}) = \eta_r(f)\chi_{q''}(r)\chi_{q''}(-1).$$

Now by Sect. 3 of Li [32] or Proposition 14.20 of [25] we have since $\chi_{q/q''}$ is real that

$$\begin{aligned}\eta_{rq^2}(f \otimes \chi_q) &= \chi_{q/q''}(-rq''^2)\eta_{rq''^2}(f \otimes \chi_{q''}) \\ &= \chi_{q/q''}(-rq''^2)\chi_{q''}(r)\chi_{q''}(-1)\eta_r(f) \\ &= \chi_q(-r)\eta_r(f).\end{aligned}$$

Note $\eta_r(f)$ is the eigenvalue of the Atkin–Lehner operator W_r on f. In our case, f is of trivial central character and square-free conductor rq'. In this case one can compute for each $p \mid rq'$ that

$$\eta_p(f) = -p^{1/2}\lambda_f(p), \tag{63}$$

for which see the proof of Theorem 2.1 of [3].

In summary, if we let ϵ_g denote the root number of a newform g, this shows

$$\begin{aligned} \epsilon_{f\otimes\chi_q} &= i^{\kappa}\chi_q(-r)\mu(r)r^{1/2}\lambda_f(r) \\ &= \chi_q(-r)\mu(q')q'^{1/2}\lambda_f(q')\epsilon_f, \end{aligned} \tag{64}$$

where recall $\lambda_f(n)$ is normalized to be bounded by the divisor function of n.

Now let

$$\omega_f := c_\kappa \frac{1}{\nu(\tilde{q}/q')\rho_f(\tilde{q}/q')\langle f, f\rangle_{rq'}}, \tag{65}$$

where ρ_f was defined in (29), and ω_f in particular satisfies

$$\omega_f = (rq)^{-1+o(1)},$$

since by [26, Lemma 2.5] [22,23] we have

$$\langle f, f\rangle_{rq'} = (rq')^{1+o(1)}.$$

Note that with these weights we have

$$\sum_{\substack{f\in H_\kappa^*(rq') \\ q'\mid\tilde{q}}} \omega_f\lambda_f(m)\lambda_f(n) = \widetilde{\Delta}_{r,\tilde{q}}(m,n), \tag{66}$$

where recall $\widetilde{\Delta}_{r,q}(m,n)$ was defined in (41).

8.2 Approximate functional equation

Recall that our goal is the bound (1), which we write as

$$\mathcal{M}(r,q) := \sum_{\substack{f\in H_\kappa^*(rq') \\ q'\mid\tilde{q}}} \omega_f L(1/2, f\otimes\chi_q)^3 \ll (qr)^{\varepsilon}. \tag{67}$$

We have for $\mathrm{Re}(s) > 1$ that $L(s, f \otimes \chi_q) = \sum_{n\geq 1} \frac{\chi_q(n)\lambda_f(n)}{n^s}$ and

$$L(s, f\otimes\chi_q)^2 = \sum_{m\geq 1}\sum_{n\geq 1}\frac{\lambda_f(m)\lambda_f(n)\chi_q(mn)}{(mn)^s} = \sum_{m\geq 1}\sum_{n\geq 1}\frac{\chi_q(mn)}{(mn)^s}\sum_{\substack{d|(m,n)\\(d,qr)=1}}\lambda_f\left(\frac{mn}{d^2}\right)$$
$$= \sum_{(d,qr)=1}\frac{1}{d^{2s}}\sum_{m\geq 1}\sum_{n\geq 1}\frac{\lambda_f(mn)\chi_q(mn)}{(mn)^s} = \sum_{(d,qr)=1}\sum_{n\geq 1}\frac{\tau(n)\chi_q(n)\lambda_f(n)}{(d^2n)^s}.$$

Then we have by standard approximate functional equations that

$$L(1/2, f\otimes\chi_q) = (1+\epsilon_{f\otimes\chi_q})\sum_{n\geq 1}\frac{\lambda_f(n)\chi_q(n)}{n^{1/2}}V_1\left(\frac{n}{q\sqrt{r}}\right),$$

and

$$L(1/2, f\otimes\chi_q)^2 = 2\sum_{(d,qr)=1}\sum_{m\geq 1}\frac{\lambda_f(m)\tau(m)\chi_q(m)}{d\sqrt{m}}V_2\left(\frac{d^2m}{rq^2}\right),$$

where V_1 and V_2 are certain bounded smooth functions of rapid decay (see (81), (82) below for formulas). Therefore,

$$\mathcal{M}(r,q) = 2\sum_{\substack{f\in H_\kappa^*(R)\\ r|R|r\tilde{q}}}\omega_f(1+\epsilon_{f\otimes\chi_q})\sum_{\substack{d,m,n\geq 1\\(d,qr)=1}}\frac{\lambda_f(m)\tau(m)\lambda_f(n)\chi_q(mn)}{d\sqrt{mn}}$$
$$\times V_1\left(\frac{n}{q\sqrt{r}}\right)V_2\left(\frac{d^2m}{q^2r}\right). \tag{68}$$

In (68), we may replace $(1+\epsilon_{f\otimes\chi_q})$ by 2, because if $\epsilon_{f\otimes\chi_q} = -1$, then the other factor $L(1/2, f\otimes\chi_q)^2$ vanishes anyway. Using this and (66), we derive

$$\mathcal{M}(r,q) = 4\sum_{(d,qr)=1}\frac{1}{d}\sum_{n\geq 1}\sum_{m\geq 1}\frac{\tau(m)\chi_q(mn)}{\sqrt{mn}}V_1\left(\frac{n}{q\sqrt{r}}\right)V_2\left(\frac{d^2m}{q^2r}\right)\widetilde{\Delta}_{r,\tilde{q}}(m,n). \tag{69}$$

The contribution from $m \gg r^{1+\varepsilon}q^{2+\varepsilon}d^{-2}$ or $n \gg r^{1/2+\varepsilon}q^{1+\varepsilon}$ is very small by trivial bounds.

8.3 Exercises with arithmetical functions

Equation (69) gives

$$\mathcal{M}(r,q) = 4\sum_{(d,qr)=1}\frac{1}{d}\mathcal{B}_{r,q}(\alpha,\beta) + O((qr)^{-A}), \tag{70}$$

where

$$\mathcal{B}_{r,q}(\alpha,\beta)=\sum_{m,n\geq 1}\alpha_m\beta_n\widetilde{\Delta}_{r,\tilde{q}}(m,n),$$

with

$$\alpha_m=\frac{\tau(m)\chi_q(m)}{\sqrt{m}}V_2\left(\frac{d^2m}{rq^2}\right),\quad \beta_n=\frac{\chi_q(n)}{\sqrt{n}}V_1\left(\frac{n}{q\sqrt{r}}\right). \tag{71}$$

Now we work in more generality than what is required.

Proposition 1 *Let α and β be two sequences of complex numbers of rapid decay, and let Y be some large power of qr. Then*

$$\begin{aligned}\mathcal{B}_{r,q}(\alpha,\beta)=&\sum_{LR=r}\frac{\mu(L)}{\nu(L)}\sum_{\substack{\ell\mid L^\infty\\ \ell\leq Y}}\frac{\ell}{\nu(\ell)^2}\sum_{d_1,d_2\mid\ell}c_\ell(d_1)c_\ell(d_2)\sum_{\substack{u\mid L\\ v\mid L}}\frac{uv}{(u,v)}\frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}\\ &\times\sum_{\substack{a\mid\frac{u}{(u,v)}\\ b\mid\frac{v}{(u,v)}}}\sum_{\substack{e_1\mid d_1\\ e_2\mid d_2}}\mathcal{B}'+O(\|\alpha_m m^\varepsilon\|_1\|\beta_n n^\varepsilon\|_1(rq)^{-100}),\end{aligned} \tag{72}$$

where

$$\begin{aligned}\mathcal{B}'=&\sum_{m\geq 1}\sum_{n\geq 1}\alpha_{au\frac{e_1}{\left(e_1,\frac{u}{a(u,v)}\right)}m}\beta_{bv\frac{e_2}{\left(e_2,\frac{v}{b(u,v)}\right)}n}\\ &\times\Delta_{R\tilde{q}}\left(\frac{d_1}{e_1}\frac{u}{(u,ae_1(u,v))}m,\frac{d_2}{e_2}\frac{v}{(v,be_2(u,v))}n\right)\end{aligned} \tag{73}$$

Proof Using (62), and pulling the sums over m and n to the inside, we obtain

$$\begin{aligned}\mathcal{B}_{r,q}(\alpha,\beta)=&\sum_{LR=r}\frac{\mu(L)}{\nu(L)}\sum_{\substack{\ell\mid L^\infty\\ \ell\leq Y}}\frac{\ell}{\nu(\ell)^2}\sum_{d_1,d_2\mid\ell}c_\ell(d_1)c_\ell(d_2)\sum_{\substack{u\mid L\\ v\mid L}}\frac{uv}{(u,v)}\frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)}\\ &\times\sum_{\substack{a\mid\frac{u}{(u,v)}\\ b\mid\frac{v}{(u,v)}}}\sum_{\substack{e_1\mid d_1\\ e_2\mid d_2}}\sum_{\substack{m\equiv 0\left(\mathrm{mod}\ au\frac{e_1}{\left(e_1,\frac{u}{a(u,v)}\right)}\right)\\ n\equiv 0\left(\mathrm{mod}\ bv\frac{e_2}{\left(e_2,\frac{v}{b(u,v)}\right)}\right)}}\alpha_m\beta_n\Delta_{R\tilde{q}}\left(\frac{md_1}{a^2e_1^2(u,v)},\frac{nd_2}{b^2e_2^2(u,v)}\right)\\ &+O\left(\sum_{m\geq 1}\sum_{n\geq 1}|\alpha_m m^\varepsilon||\beta_n n^\varepsilon|(rq)^{1+\varepsilon}Y^{-2\gamma_0+\varepsilon}\right),\end{aligned}$$

where we have used the following elementary observations: we have

$$\begin{cases} u|m \\ a|\frac{m}{u} \end{cases} \Leftrightarrow m \equiv 0 \pmod{au}$$

and for any integers a, b, x we have $ax \equiv 0 \pmod{b}$ if and only if $x \equiv 0 \pmod{\frac{b}{(a,b)}}$ so that

$$\begin{aligned} \frac{m}{a^2(u,v)} &\equiv 0 \pmod{e_1} \\ &\Leftrightarrow \frac{m}{au}\frac{u}{a(u,v)} \equiv 0 \pmod{e_1} \\ &\Leftrightarrow \frac{m}{au} \equiv 0 \left(\mathrm{mod}\ \frac{e_1}{\left(e_1, \frac{u}{a(u,v)}\right)}\right) \\ &\Leftrightarrow m \equiv 0 \left(\mathrm{mod}\ au\frac{e_1}{\left(e_1, \frac{u}{a(u,v)}\right)}\right). \end{aligned}$$

We now make the change of variables

$$m \mapsto au\frac{e_1}{\left(e_1, \frac{u}{a(u,v)}\right)}m \quad n \mapsto bv\frac{e_2}{\left(e_2, \frac{v}{b(u,v)}\right)}n,$$

which gives the desired formula. □

Continuing with our more general set-up, let $\gamma_1, \gamma_2, \delta_1, \delta_2$ be positive integers that divide L^∞, and set

$$\mathcal{B}'_{\gamma_1,\gamma_2,\delta_1,\delta_2} = \sum_{m,n\geq 1} \alpha_{\gamma_1 m}\beta_{\gamma_2 n}\Delta_{R\tilde{q}}(\delta_1 m, \delta_2 n). \tag{74}$$

In our application of interest, we have

$$\gamma_1 = au\frac{e_1}{\left(e_1, \frac{u}{a(u,v)}\right)}, \qquad \gamma_2 = bv\frac{e_2}{\left(e_2, \frac{v}{b(u,v)}\right)}, \tag{75}$$

and

$$\delta_1 = \frac{d_1}{e_1}\frac{u}{(u, ae_1(u,v))}, \qquad \delta_2 = \frac{d_2}{e_2}\frac{v}{(v, be_2(u,v))}. \tag{76}$$

We now use the α and β specific to our situation. In anticipation of some future maneuvers, we use a Hecke relation on the divisor function implicit in α, namely

$\tau(\gamma m) = \sum_{h|(\gamma,m)} \tau(\gamma/h)\tau(m/h)\mu(h)$. This gives (for an arbitrary function f such that the sums converge absolutely)

$$\begin{aligned}\sum_{m\geq 1} \alpha_{\gamma m} f(m) &= \sum_{m\geq 1} \frac{\tau(\gamma m)\chi_q(\gamma m)}{\sqrt{\gamma m}} V_2\left(\frac{d^2\gamma m}{q^2 r}\right) f(m)\\ &= \frac{\chi_q(\gamma)}{\sqrt{\gamma}} \sum_{h|\gamma} \frac{\tau\left(\frac{\gamma}{h}\right)\mu(h)\chi_q(h)}{\sqrt{h}} \sum_{m\geq 1} \frac{\chi_q(m)\tau(m)}{\sqrt{m}} V_2\left(\frac{d^2\gamma hm}{q^2 r}\right) f(hm).\end{aligned}$$

With this, and by inserting the definition of β, we have

$$\mathcal{B}'_{\gamma_1,\gamma_2,\delta_1,\delta_2} = \frac{\chi_q(\gamma_1\gamma_2)}{\sqrt{\gamma_1\gamma_2}} \sum_{\gamma_3|\gamma_1} \frac{\tau(\frac{\gamma_1}{\gamma_3})\mu(\gamma_3)\chi_q(\gamma_3)}{\sqrt{\gamma_3}} \mathcal{B}'', \tag{77}$$

where

$$\mathcal{B}'' = \sum_{m,n\geq 1} \frac{\chi_q(mn)\tau(m)}{\sqrt{mn}} V_1\left(\frac{\gamma_2 n}{q\sqrt{r}}\right) V_2\left(\frac{d^2\gamma_1\gamma_3 m}{q^2 r}\right) \Delta_{R\tilde{q}}(\delta_1\gamma_3 m, \delta_2 n). \tag{78}$$

Applying the Petersson formula (11), we obtain

$$\mathcal{B}'' = \mathcal{D}'' + 2\pi i^{-\kappa}\mathcal{S},$$

where

$$\mathcal{D}'' = \sum_{\delta_1\gamma_3 m = \delta_2 n} \frac{\chi_q(mn)\tau(m)}{\sqrt{mn}} V_1\left(\frac{\gamma_2 n}{q\sqrt{r}}\right) V_2\left(\frac{d^2\gamma_1\gamma_3 m}{q^2 r}\right), \tag{79}$$

and

$$\begin{aligned}\mathcal{S} = \sum_{m,n\geq 1} &\frac{\chi_q(mn)\tau(m)}{\sqrt{mn}} V_1\left(\frac{\gamma_2 n}{q\sqrt{r}}\right) V_2\left(\frac{d^2\gamma_1\gamma_3 m}{q^2 r}\right)\\ &\times \sum_{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)} \frac{S(Am, Bn; c)}{c} J_{\kappa-1}\left(\frac{4\pi\sqrt{ABmn}}{c}\right),\end{aligned} \tag{80}$$

with

$$A = \delta_1\gamma_3, \quad B = \delta_2.$$

According to this, we write $\mathcal{B}' = \mathcal{D}' + \mathcal{B}'_{\mathcal{S}}$, and similarly, $\mathcal{B} = \mathcal{D} + \mathcal{B}_{\mathcal{S}}$. It may be helpful to record that $c \equiv 0 \pmod{\tilde{q}R}$, that $(qR, L) = 1$, and that $AB|L^\infty$, so that $(AB, qR) = 1$.

The main technical result proved in the rest of the paper is the following

Proposition 2 *With α and β as above, we have*

$$\mathcal{S} \ll \left(\frac{\sqrt{AB}}{\sqrt{R}} + \frac{r^{3/4}}{\sqrt{q}R} \right) (qr)^{\varepsilon}.$$

From Proposition 2, we deduce bounds on $\mathcal{B}'_{\mathcal{S}}$, then $\mathcal{B}_{\mathcal{S}}$. We have

$$\mathcal{B}'_{\mathcal{S}} \ll \frac{1}{\sqrt{\gamma_1\gamma_2}} \sum_{\gamma_3|\gamma_1} \frac{1}{\sqrt{\gamma_3}} \left(\frac{\sqrt{\delta_1\delta_2\gamma_3}}{\sqrt{R}} + \frac{r^{3/4}}{q^{1/2}R} \right) (qr)^{\varepsilon} \ll \left(\frac{\sqrt{\delta_1\delta_2}}{\sqrt{R}} + \frac{r^{3/4}}{q^{1/2}R} \right) \frac{(qr)^{\varepsilon}}{\sqrt{\gamma_1\gamma_2}}.$$

Therefore, we get the following bound on $\mathcal{B}_{\mathcal{S}}$:

$$\begin{aligned} \mathcal{B}_{\mathcal{S}} \ll (qr)^{\varepsilon} \sum_{LR=r} \frac{1}{\nu(L)} \sum_{\substack{\ell | L^{\infty} \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2|\ell} c_{\ell}(d_1)c_{\ell}(d_2) \\ \times \sum_{\substack{u|L \\ v|L}} \frac{uv}{(u,v)} \frac{\left|\mu\left(\frac{uv}{(u,v)^2}\right)\right|}{\nu\left(\frac{uv}{(u,v)^2}\right)} \sum_{\substack{a|\frac{u}{(u,v)} \\ b|\frac{v}{(u,v)}}} \sum_{\substack{e_1|d_1 \\ e_2|d_2}} \left(\frac{\sqrt{\delta_1\delta_2}}{\sqrt{R}} + \frac{r^{3/4}}{q^{1/2}R} \right) \frac{1}{\sqrt{\gamma_1\gamma_2}}. \end{aligned}$$

Note that

$$\frac{\delta_1}{\gamma_1} = \frac{d_1}{e_1} \frac{u}{(u, ae_1(u,v))} \frac{\left(e_1, \frac{u}{a(u,v)}\right)}{aue_1} = \frac{d_1}{e_1^2 a^2 (u,v)},$$

and so by symmetry

$$\frac{\delta_2}{\gamma_2} = \frac{d_2}{e_2^2 b^2 (u,v)},$$

and thus

$$\left(\frac{\delta_1\delta_2}{\gamma_1\gamma_2} \right)^{1/2} = \frac{(d_1d_2)^{1/2}}{e_1e_2ab(u,v)}.$$

We also use $\frac{1}{\sqrt{\gamma_1\gamma_2}} \le \frac{1}{\sqrt{uv}} \le \frac{1}{(u,v)}$. With these observations, we have

$$\sum_{\substack{a|\frac{u}{(u,v)} \\ b|\frac{v}{(u,v)}}} \sum_{\substack{e_1|d_1 \\ e_2|d_2}} \left(\frac{\sqrt{\delta_1\delta_2}}{\sqrt{R}} + \frac{r^{3/4}}{q^{1/2}R} \right) \frac{1}{\sqrt{\gamma_1\gamma_2}} \ll \frac{(qr)^{\varepsilon}}{(u,v)} \left(\frac{(d_1d_2)^{1/2}}{R^{1/2}} + \frac{r^{3/4}}{q^{1/2}R} \right).$$

The inner sum over u and v gives a divisor bound, so now we get

$$\mathcal{B}_{\mathcal{S}} \ll (qr)^{\varepsilon} \sum_{LR=r} \frac{1}{\nu(L)} \sum_{\substack{\ell | L^{\infty} \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2|\ell} c_{\ell}(d_1)c_{\ell}(d_2) \left(\frac{(d_1d_2)^{1/2}}{R^{1/2}} + \frac{r^{3/4}}{q^{1/2}R} \right).$$

By Lemma 4, we bound the second part of this sum by

$$(qr)^{\varepsilon}\frac{r^{3/4}}{q^{1/2}}\sum_{LR=r}\frac{1}{\nu(L)R}\sum_{\substack{\ell\mid L^{\infty}\\ \ell\le Y}}\frac{\ell}{\nu(\ell)^2}\sum_{d_1,d_2\mid\ell}c_\ell(d_1)c_\ell(d_2)$$
$$\ll (qr)^{\varepsilon}\frac{r^{3/4}}{q^{1/2}}\sum_{LR=r}\frac{1}{\nu(L)R}\sum_{\substack{\ell\mid L^{\infty}\\ \ell\le Y}}\ell^{-2\gamma_0}\ll q^{-1/2+\varepsilon}r^{-1/4+\varepsilon}.$$

Recalling the definition (53) and using Lemma 3 we have

$$\mathcal{B}_{\mathcal{S}}\ll (qr)^{\varepsilon}\left(\sum_{LR=r}\frac{S(L,Y)}{\nu(L)R^{1/2}}+r^{-1/4}q^{-1/2}\right)\ll (qr)^{\varepsilon}\left(r^{-1/2}+r^{-1/4}q^{-1/2}\right).$$

Finally, we have from (70) that

$$\mathcal{M}(r,q)=\mathcal{M}_0(r,q)+O\left((qr)^{\varepsilon}(r^{-1/4}q^{-1/2}+r^{-1/2})\right),$$

where $\mathcal{M}_0(r,q)$ is the contribution to $\mathcal{M}(r,q)$ from the diagonal term $\mathcal{D}$. It is easy to see that $\mathcal{M}_0(r,q)\ll (rq)^{\varepsilon}$, following the proof of the bounds on $\mathcal{B}_S$.

We summarize this discussion with

Corollary 3 *Proposition 2 implies Theorem 1.*

This is appealing because it lets us reduce the number of variables to consider from this point onward.

8.4 Diagonal terms

In this section, we evaluate $\mathcal{M}_0(r,q)$ which along with Proposition 2 leads to (5).

The functions V_1 and V_2, are given explicitly by

$$V_i(y)=\frac{1}{2\pi i}\int_{(2)}W_i(u_i)y^{-u_i}\,du_i, \tag{81}$$

where

$$W_1(u)=(2\pi)^{-u}\frac{\Gamma\left(u+\frac{\kappa}{2}\right)}{\Gamma\left(\frac{\kappa}{2}\right)u},\quad W_2(u)=(2\pi)^{-2u}\frac{\Gamma\left(u+\frac{\kappa}{2}\right)^2}{\Gamma\left(\frac{\kappa}{2}\right)^2u}. \tag{82}$$

Then recalling (70), (72), (77), and (79) we have that

$$\mathcal{M}_0(r,q)=\int_{(1)}\int_{(1)}W_1(u_1)W_2(u_2)(q^2r)^{\frac{u_1}{2}+u_2}\zeta(1+2u_2)\zeta(1+u_1+u_2)^2$$
$$\times F_{r,q,Y}(u_1,u_2)\frac{4du_1du_2}{(2\pi i)^2},$$

where

$$F_{r,q,Y}(u_1,u_2) = \sum_{LR=r} \frac{\mu(L)}{\nu(L)} \sum_{\ell \mid L^\infty \ell \le Y} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2)$$
$$\times \sum_{\substack{u \mid L \\ v \mid L}} \frac{uv}{(u,v)} \frac{\mu\left(\frac{uv}{(u,v)^2}\right)}{\nu\left(\frac{uv}{(u,v)^2}\right)} \sum_{\substack{a \mid \frac{u}{(u,v)} \\ b \mid \frac{v}{(u,v)}}} \sum_{\substack{e_1 \mid d_1 \\ e_2 \mid d_2}} \frac{\chi_q(\gamma_1\gamma_2)}{\gamma_1^{1/2+u_2}\gamma_2^{1/2+u_1}}$$
$$\times \sum_{\gamma_3 \mid \gamma_1} \frac{\tau\left(\frac{\gamma_1}{\gamma_3}\right)\mu(\gamma_3)\chi_q(\gamma_3)}{\gamma_3^{1/2+u_2}}$$
$$\times \frac{\prod_{p \mid qr}(1-p^{-1-2u_2})}{\zeta(1+u_1+u_2)^2} \sum_{\delta_1\gamma_3 m = \delta_2 n} \frac{\chi_q(mn)\tau(m)}{m^{1/2+u_2}n^{1/2+u_1}},$$

and where recall (75) and (76) for the definitions of $\gamma_1, \gamma_2, \delta_1, \delta_2$, which depend on $a, b, u, v, e_1, e_2, d_1, d_2$.

Our plan is to shift the contours past the poles. We claim $F_{r,q,Y}(u_1,u_2)$ is holomorphic in the region $\mathrm{Re}(u_i) = \sigma_i \ge -1/2$, for $i = 1, 2$, and satisfies the bound

$$|F_{r,q,Y}(u_1,u_2)| \ll (qr)^\varepsilon. \tag{83}$$

Proof By a simple argument with Euler factors, it is not hard to see that we have the bound

$$\zeta(1+u_1+u_2)^{-2} \sum_{\delta_1\gamma_3 m = \delta_2 n} \frac{\chi_q(mn)\tau(m)}{m^{1/2+u_2}n^{1/2+u_1}} \ll (qr)^\varepsilon, \tag{84}$$

and that the left hand side of (84), and hence $F_{r,q,Y}$, is holomorphic in the desired region.

Using divisor-type bounds on the inner sums, we have

$$|F_{r,q,Y}(u_1,u_2)| \ll (qr)^\varepsilon \sum_{LR=r} \frac{1}{\nu(L)} \sum_{\substack{\ell \mid L^\infty \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2).$$

By Lemma 4, we have

$$\sum_{\substack{\ell \mid L^\infty \\ \ell \le Y}} \frac{\ell}{\nu(\ell)^2} \sum_{d_1,d_2 \mid \ell} c_\ell(d_1)c_\ell(d_2) \ll L^\varepsilon,$$

and hence (83) follows. □

The proof given above, combined with Corollary 2, shows that

$$F_{r,q,Y} = \lim_{Y \to \infty} F_{r,q,Y} + O((qr)^\varepsilon Y^{-2\gamma_0+2\varepsilon}),$$

so for the rest of the calculation of $\mathcal{M}_0(r,q)$ we take $Y=\infty$, and define $F_{r,q}=\lim_{Y\to\infty}F_{r,q,Y}$.

Rather than attempting to obtain the strongest error term, we take the easiest path that gives some power saving. We begin by taking $\sigma_1=1/2+\varepsilon$, and $\sigma_2=1/2$. Next we shift u_2 to the line $\sigma_2=-1/2$, crossing a double pole at $u_2=0$ only. On the new line, we have

$$\int_{(\sigma_1)}\int_{(\sigma_2)}|W_1(u_1)W_2(u_2)(q^2r)^{u_1/2+u_2}\zeta(1+u_1+u_2)^2\zeta(1+2u_2)\times F_{r,q}(u_1,u_2)||du_1du_2|\ll(rq^2)^{-1/4+\varepsilon}.$$

Some thought shows that

$$\begin{aligned}&\mathrm{Res}_{u_2=0}W_2(u_2)(q^2r)^{u_2}\zeta(1+u_1+u_2)^2\zeta(1+2u_2)F_{r,q}(u_1,u_2)\\&=F_{r,q}(u_1,0)P_1(\log q^2r)\zeta^2(1+u_1)+c\zeta'(1+u_1)\zeta(1+u_1)F_{r,q}(u_1,0)\\&\quad+c'\zeta^2(1+u_1)F_{r,q}^{(0,1)}(u_1,0),\end{aligned}\tag{85}$$

where c,c' are constants and P_1 is a degree 1 polynomial.

The residue is now a single integral over u_1, and we shift this contour to $\sigma_1=-1/2+\varepsilon$. The new integral is bounded by $(q^2r)^{-1/4+\varepsilon}$, again. The residue at $u_1=0$ takes the form

$$R_{r,q}:=\sum_{\substack{0\le i\le 2\\0\le j\le 1}}P_{i,j}(\log q^2r)F_{r,q}^{(2-i,1-j)}(0,0),\tag{86}$$

where $P_{i,j}$ is a polynomial of degree $\le i+j$.

Gathering this discussion together, we have shown

$$\mathcal{M}_0(r,q)=R_{r,q}+O((q^2r)^{-1/4+\varepsilon}).$$

It would be better to study the main terms in the style of [13] using shifts, which for the sake of brevity we leave for another occasion.

8.5 Dyadic subdivisions

We return to estimating $\mathcal{S}$ defined by (80). Next, we open the divisor function $\tau(m)=\sum_{n_1n_2=m}1$ and apply a dyadic partition of unity to the sums over $n_1,n_2,n=n_3$, and c. This gives

$$\mathcal{S}=\sum_{N_1,N_2,N_3,C}\frac{1}{(N_1N_2N_3)^{1/2}C}\mathcal{S}_{N_1,N_2,N_3,C}+O((qr)^{-10}),$$

where N_1, N_2, N_3, C run over dyadic numbers and where

$$\mathcal{S}_{N_1,N_2,N_3,C} = \sum_{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)} w_C(c) \sum_{n_1,n_2,n_3\geq 1} \chi_q(n_1n_2n_3) \times S(An_1n_2, Bn_3; c) J_{\kappa-1}\left(\frac{4\pi\sqrt{ABn_1n_2n_3}}{c}\right) w_{N_1,N_2,N_3}(n_1,n_2,n_3). \tag{87}$$

Here the weight functions $w_C(x)$ and $w_{N_1,N_2,N_3}(x_1,x_2,x_3)$ satisfy

$$w_C^{(j)}(x) \ll x^{-j}, \quad w_{N_1,N_2,N_3}^{(j_1,j_2,j_3)}(x_1,x_2,x_3) \ll x_1^{-j_1}x_2^{-j_2}x_3^{-j_3},$$

and are supported on $x \asymp C$, $x_i \asymp N_i$, $i = 1, 2, 3$.

By the Weil bound, and using $J_{\kappa-1}(x) \ll x$, the contribution to $\mathcal{S}$ from $c \geq C$ is

$$\ll \frac{(qr)^\varepsilon}{\sqrt{N_1N_2N_3}C}\left(\frac{\sqrt{ABN_1N_2N_3}}{C}\right)\frac{C^{3/2}}{qR}N_1N_2N_3 = \frac{\sqrt{AB}N_1N_2N_3}{C^{1/2}qR}(qr)^\varepsilon.$$

This is satisfactory for Proposition 2 for $C \gg \frac{(N_1N_2N_3)^2}{q^2R}$. Thus we may restrict the variables by

$$qR \ll C \ll \frac{(N_1N_2N_3)^2}{q^2R}, \quad N_1N_2 \ll \frac{(q^2r)^{1+\varepsilon}}{d^2\gamma_1\gamma_3}, \quad N_3 \ll \frac{(qr^{1/2})^{1+\varepsilon}}{\gamma_2}. \tag{88}$$

Let us also write (87) as

$$\mathcal{S}_{N_1,N_2,N_3,C} = \sum_{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)} w_C(c)\mathcal{S}'_{N_1,N_2,N_3,c}. \tag{89}$$

8.6 Poisson summation

Let $[c, q] = \mathrm{lcm}(c, q)$. We have

$$\mathcal{S}'_{N_1,N_2,N_3,c} = \sum_{m_1,m_2,m_3\in\mathbb{Z}} G_{A,B}(m_1,m_2,m_3;c)K(m_1,m_2,m_3;c), \tag{90}$$

where

$$G_{A,B}(m_1,m_2,m_3;c) = \frac{1}{[c,q]^3}\sum_{x_1,x_2,x_3 (\mathrm{mod}\ [c,q])} \chi_q(x_1x_2x_3) \times S(Ax_1x_2, Bx_3; c) e\left(\frac{x_1m_1 + x_2m_2 + x_3m_3}{[c,q]}\right), \tag{91}$$

and

$$K(m_1, m_2, m_3; c) = \int_{\mathbb{R}^3} J_{\kappa-1}\left(\frac{4\pi\sqrt{ABt_1t_2t_3}}{c}\right) e\left(\frac{-m_1t_1 - m_2t_2 - mt_3}{[c,q]}\right) \times w_{N_1,N_2,N_3}(t_1, t_2, t_3)dt_1dt_2dt_3. \tag{92}$$

When $A = B = 1$ and q is odd and square-free, this is precisely as in [14] (though the reader should be aware of our slightly different normalization of G by $[c, q]^{-3}$), so this appears at first glance to be a fairly minor generalization of their work, however the calculations become rather intricate.

9 Arithmetic part

Let $(\epsilon_1, \epsilon_2, \epsilon_3) \in \{\pm 1\}^3$, $\delta = 1, 2, 4$, or 8, and write q_e for the even part of q. Let

$$Z_{\delta,R,q}^{(\epsilon_1,\epsilon_2,\epsilon_3)}(s_1, s_2, s_3, s_4) = \sum_{\substack{\epsilon_i m_i \geq 1 \\ i=1,2,3}} \sum_{\substack{c \equiv 0 (\mathrm{mod}\ R\tilde{q}) \\ (c,q_e)=\delta}} \frac{cqG_{A,B}(m_1, m_2, m_3; c)e_{AB[c,q]^3/c^2}(-m_1m_2m_3)}{|m_1|^{s_1}|m_2|^{s_2}|m_3|^{s_3}\left(\frac{c}{R\tilde{q}}\right)^{s_4}}. \tag{93}$$

One of the key ingredients of the Conrey-Iwaniec method (when $A = B = 1$ and q is odd) is that the additive character $e_{AB[c,q]^3/c^2}(-m_1m_2m_3)$ nicely combines with $G(m_1, m_2, m_3; c)$, allowing for an efficient decomposition into multiplicative characters.

To avoid over-burdening the already burdened notation we only give proofs in the case $(\epsilon_1, \epsilon_2, \epsilon_3) = (1, 1, 1)$ and denote this case simply $Z_{\delta,R,q}$, the other cases being treated similarly. Note that we have $(AB) \ll (qr)^L$ for some fixed but possibly large $L > 0$ (see Proposition 1) so that $(AB)^\varepsilon \ll (qr)^{\varepsilon'}$.

The main goal of this section is the following proposition.

Proposition 3 *For each choice of $(\epsilon_1, \epsilon_2, \epsilon_3)$ and δ there is a decomposition $Z_{\delta,R,q} = Z_{\delta,R,q}^{(\epsilon_1,\epsilon_2,\epsilon_3)}(s_1, s_2, s_3, s_4) = Z_0 + Z'$, where Z_0 and Z' have the following properties. Here Z_0 is analytic in $Re(s_i) \geq 1 + \sigma$ for $i = 1, 2, 3, 4$, $\sigma > 0$ and in this region it satisfies the bound*

$$Z_0 \ll_{\sigma,\varepsilon} \frac{(qr)^\varepsilon}{AB}. \tag{94}$$

The function Z' is analytic for $Re(s_i) \geq \frac{1}{2} + \sigma$ for $i = 1, 2, 3, 4$, any $\sigma > 0$, and in this region satisfies the bound

$$|Z'| \ll_{\sigma,\varepsilon} q^{3/2}(AB)^{1/2}(qr)^\varepsilon \prod_{j=1}^{4}(1 + |s_j|)^{1/4+\varepsilon}. \tag{95}$$

Moreover, if $s_j = 1/2 + \varepsilon + i(y_j + t)$ *for* $j = 1, 2, 3$, $s_4 = 1/2 + \varepsilon + i(y_4 - t)$, *and* $y_j \ll (qr)^\varepsilon$ *for* $j = 1, 2, 3, 4$, *then we have*

$$\int_{|t|\le T} |Z'(1/2+\varepsilon+i(y_1+t), 1/2+\varepsilon+i(y_2+t), 1/2+\varepsilon+i(y_3+t), 1/2+\varepsilon+i(y_4-t))|dt \ll q^{3/2}(AB)^{1/2}T(qrT)^\varepsilon. \tag{96}$$

We begin by reducing the evaluation of $G_{A,B}$ into cases. First, write $c = c_1c_2$ with $c_2|(AB)^\infty$, and $(c_1, AB) = 1$. As $r = RL$ is square-free and $(r, q) = 1$ we have $(qR, AB) = 1$, hence $(qR, c_2) = 1$. By a calculation with the Chinese remainder theorem, we have

$$G_{A,B}(m_1, m_2, m_3; c_1c_2) = \chi_q(AB)G_{1,1}\left(m_1, m_2, \overline{ABc_2}m_3; c_1\right) \times G_{A,B}\left(m_1, m_2, \overline{[c_1, q]}^3 c_1^2 m_3; c_2\right), \tag{97}$$

where in the definition of $G_{A,B}(m_1, m_2, m_3; c_2)$ for $c_2 \mid (AB)^\infty$ we implicitly take $q = 1$. Write $q = q_oq_e$ where q_o is odd and $q_e \in \{1, 4, 8\}$. We further decompose c_1 by $c_1 = c_oc_e$ where c_o is odd and c_e is a power of 2. Another short calculation with the Chinese remainder theorem shows

$$G_{1,1}(\ell_1, \ell_2, \ell_3; c_oc_e) = G_{1,1}(\ell_1, \ell_2, \overline{[c_e, q_e]}^3 c_e^2\ell_3; c_o)G_{1,1}(\ell_1, \ell_2, \overline{c_o}\ell_3; c_e). \tag{98}$$

Next we evaluate the three types of $G_{A,B}$ sums in a form most relevant for our further calculations. The case with modulus c_o was derived by [14, §10]. Following the notation found in [14], write

$$c_o = q_os_o.$$

Lemma 5 (Conrey and Iwaniec) *We have for* $q_o, s_o \in \mathbb{N}$ *with* $(q_o, 2) = 1$ *that*

$$c_oq_oe\left(\frac{-a_1a_2a_3}{c_o}\right)G_{1,1}(a_1, a_2, a_3; c_o) = \sum_{\substack{D_1D_2hk=q_o \\ h=(q_o,s_o) \\ k=(a_1a_2a_3,q_o) \\ (h,a_1a_2)=1}} \frac{1}{\varphi(D_2)} \times \sum_{\psi \,(\mathrm{mod}\ D_2)} g_{D_1,D_2,h,k,\psi}\psi(a_1a_2a_3)\overline{\psi}(s_o) R_k(a_1)R_k(a_2)R_k(a_3), \tag{99}$$

where $R_k(n) = S(n, 0; k)$ *is the Ramanujan sum, and* g *is some function satisfying*

$$|g_{D_1,D_2,h,k,\psi}| \ll D_2^{3/2+\varepsilon}, \tag{100}$$

and where in addition we must have $(a_3, s_o) = 1$, *otherwise* G *vanishes. In case* ψ *is principal, then* $|g_{D_1,D_2,h,k,\psi}| \ll D_2^\varepsilon$.

Conrey and Iwaniec in fact give a more precise formula that we describe within the proof.

Next we evaluate the case with modulus $c_2|(AB)^\infty$.

Lemma 6 *Suppose $c_2|(AB)^\infty$. Suppose that $a_1a_2a_3 \neq 0$ and write $a_i = u_iv_i$ where $(u_i, AB) = 1$ and $v_i|(AB)^\infty$. Then*

$$\begin{aligned} &c_2 e\left(\frac{-a_1a_2a_3}{c_2AB}\right) G_{A,B}(a_1,a_2,a_3;c_2) \\ &= \delta((A,c_2)|v_1)\delta((A,c_2)|v_2)\delta((B,c_2)|v_3) \\ &\quad\times \sum_{\substack{g_1g_2|\frac{c_2}{(A,c_2)} \\ g_1=\left(\frac{v_1}{(A,c_2)},\frac{c_2}{(A,c_2)}\right) \\ g_2=\left(\frac{v_2}{(A,c_2)},\frac{c_2}{g_1(A,c_2)}\right)}} \sum_{D|\frac{c_2AB}{(c_2,A)^2(c_2,B)}} \frac{1}{\varphi(D)} \sum_{\eta \,(\mathrm{mod}\ D)} \gamma_\eta \eta(u_1u_2u_3), \end{aligned} \tag{101}$$

were $\gamma_\eta = \gamma_{v_1,v_2,g_1,g_2,c_2,A,B,D,\eta}$ is some function satisfying the bound

$$|\gamma_{v_1,v_2,g_1,g_2,c_2,A,B,D,\eta}| \ll (A,c_2)(B,c_2)D^{1/2}. \tag{102}$$

In case η is principal then with $A = (A,c_2)A'$ and $B = (B,c_2)B'$, we have

$$\frac{|\gamma_\eta|}{D} \ll (qr)^\varepsilon (A,c_2)(B,c_2)\frac{\left(\frac{v_1v_2v_3}{(A,c_2)^2(B,c_2)}, A'B'\right)}{A'B'}. \tag{103}$$

Again the point is that we get a short linear combination of multiplicative functions.

Finally, we consider the case of c_e. For this, we have

Lemma 7 *Suppose c_e is a power of 2. Suppose $a_1a_2a_3 \neq 0$ and write each $a_i = e_if_i$ where e_i is a power of 2, and f_i is odd. Then*

$$q_ec_ee\left(\frac{-a_1a_2a_3}{[c_e,q_e]^3/c_e^2}\right) G_{1,1}(a_1,a_2,a_3;c_e) = \sum_{\Delta|64} \frac{1}{\varphi(\Delta)} \sum_{\chi \,(\mathrm{mod}\ \Delta)} \mathfrak{g}_\chi \chi(f_1f_2f_3), \tag{104}$$

where $\mathfrak{g}_\chi = \mathfrak{g}_{e_1,e_2,e_3,q_e,c_e,\chi,\Delta}$ is bounded by an absolute constant.

As in the previous two cases, we have a much more precise formula for $G_{1,1}$, which we shall describe within the proof.

Proof (of Lemma 5) First we note that our $G_{1,1}$ is scaled differently from G defined by [14], precisely $G_{1,1}(a_1,a_2,a_3;c_o) = c_o^{-3}G(a_1,a_2,a_3;c_o)$, as in [14, (8.2)]. In the notation of [14], make the definitions $c_o = q_os_o$, $h = (q_o,s_o)$, $k = (a_1a_2a_3,q_o)$, $D = \frac{q_o}{hk}$. The sum $G_{1,1}$ vanishes unless $(h,a_1a_2) = 1$ and $(s_0,a_3) = 1$, in which case by [14, Lemma 10.2], we have

$$G_{1,1}(a_1, a_2, a_3; c_o) = e\left(\frac{a_1a_2a_3}{c_o}\right)\frac{h\chi_{kD}(-1)}{c_oq_o\phi(k)}R_k(a_1)R_k(a_2)R_k(a_3)H(\overline{s_ohk}a_1a_2a_3; D). \quad (105)$$

We do not need the exact formula for H, but rather the fact that it essentially depends on the variables as a block, and the decomposition into character sums. Specifically, Conrey and Iwaniec [14, (11.7), (11.9)] showed

$$H(w; D) = \sum_{D_1D_2=D}\mu(D_1)\chi_{D_1}(-1)H^*(\overline{D_1}w; D_2), \quad (106)$$

and

$$H^*(w; D_2) = \frac{1}{\varphi(D_2)}\sum_{\psi \,(\mathrm{mod}\ D_2)}\tau(\overline{\psi})g(\chi_{D_2}, \psi)\psi(w). \quad (107)$$

The crucial fact about $g(\chi_{D_2}, \psi)$ is that

$$|g(\chi_{D_2}, \psi)| \ll D_2^{1+\varepsilon}, \quad (108)$$

which requires the Riemann Hypothesis for varieties, i.e., Deligne's bound.

From here it is a matter of bookkeeping to derive (99).

In case $\psi = \psi_0$ is the principal character, then $|g(\chi_{D_2}, \psi_0)| \le d(D_2)$ (the divisor function) and $\tau(\psi_0) = \mu(D_2)$. Indeed, one may show that if $\psi = \psi_0$ is the principal character modulo an odd prime p, then $g(\chi_p, \psi_0) = 2$ if $p \equiv 1 \pmod 4$, and $= 0$ if $p \equiv 3 \pmod 4$. Furthermore, $g(\chi_q, \psi_0)$ is multiplicative in q.[1] □

Proof (of Lemma 6) We will evaluate $G_{A,B}$ in precise terms. We will not use the assumption $a_1a_2a_3 = 0$ until indicated later in the proof. Since $c_2|(AB)^\infty$ and $(q, AB) = 1$, the quadratic character is not present in the sum, and specifically we have

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{1}{c_2^3}\sideset{}{^*}\sum_{u \,(\mathrm{mod}\ c_2)}\sum_{x_1,x_2,x_3 \,(\mathrm{mod}\ c_2)} e\left(\frac{Ax_1x_2\overline{u} + Bx_3u + x_1a_1 + x_2a_2 + x_3a_3}{c_2}\right).$$

Summing over x_1, we detect the congruence $Ax_2\overline{u} \equiv -a_1 \pmod{c_2}$, while the sum over x_3 detects $Bu \equiv -a_3 \pmod{c_2}$. Therefore,

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{1}{c_2}\sideset{}{^*}\sum_{Bu\equiv -a_3 \,(\mathrm{mod}\ c_2)}\sum_{Ax_2\overline{u}\equiv -a_1 \,(\mathrm{mod}\ c_2)} e\left(\frac{x_2a_2}{c_2}\right).$$

Note that $Bu \equiv -a_3 \pmod{c_2}$ and $Ax_2\overline{u} \equiv -a_1 \pmod{c_2}$ are solvable if and only if

$$(B, c_2) = (a_3, c_2), \quad \text{and} \quad (A, c_2)|a_1. \quad (109)$$

[1] This corrects a claimed formula for $g(\chi_q, \psi_0)$ of [14, p.1212].

By symmetry, we expect that in addition that we will require $(A, c_2)|a_2$, and indeed we will recover this condition later in the analysis. From now on, we assume the conditions (109) hold, otherwise the sum is 0.

Next we make the definitions

$$A = (A, c_2)A', \quad B = (B, c_2)B', \quad c_2 = (B, c_2)c_2', \quad a_3 = (a_3, c_2)\widetilde{a_3}. \tag{110}$$

Now the congruence $Bu \equiv -a_3 \pmod{c_2}$ is equivalent to $u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_2'}$. Next write $x_2 = -\overline{A'}\frac{a_1}{(A,c_2)}u + \frac{c_2}{(A,c_2)}t$ with t running mod (A, c_2). Inserting this into the exponential sum, we obtain

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{1}{c_2} \sideset{}{^*}\sum_{\substack{u \pmod{c_2} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_2'}}} \sum_{t \pmod{(A,c_2)}} e\left(\frac{a_2(-\overline{A'}\frac{a_1}{(A,c_2)}u + \frac{c_2}{(A,c_2)}t)}{c_2}\right).$$

The sum over t vanishes unless $(A, c_2)|a_2$, in which case we obtain

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{(A, c_2)}{c_2} \sideset{}{^*}\sum_{\substack{u \pmod{c_2} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_2'}}} e\left(\frac{-\overline{A'}\frac{a_1}{(A,c_2)}\frac{a_2}{(A,c_2)}u}{\frac{c_2}{(A,c_2)}}\right).$$

To proceed further, we make some additional definitions, namely

$$g_1 = \left(\frac{a_1}{(A, c_2)}, \frac{c_2}{(A, c_2)}\right), \quad a_1 = g_1(A, c_2)\widetilde{a_1}, \quad g_2 = \left(\frac{a_2}{(A, c_2)}, \frac{c_2}{g_1(A, c_2)}\right),$$
$$a_2 = g_2(A, c_2)\widetilde{a_2}, \qquad c_2'' = \frac{c_2}{(A, c_2)g_1g_2}.$$

Thus $(\widetilde{a_1}, \frac{c_2}{g_1(A,c_2)}) = 1$, and $(A'\widetilde{a_1}\widetilde{a_2}, c_2'') = 1$, and with this notation the formula becomes

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{(A, c_2)}{c_2} \sideset{}{^*}\sum_{\substack{u \pmod{c_2} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_2'}}} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2}u}{c_2''}\right).$$

The tricky part in the analysis is that there is no apparent divisibility relationship between $c_2' = \frac{c_2}{(B,c_2)}$ and $c_2'' = \frac{c_2}{(A,c_2)g_1g_2}$, and so it is necessary to proceed by cases. Although it is not globally true that $c_2'|c_2''$, or vice versa, we may factor the moduli corresponding to which prime power of c_2' or c_2'' is larger, which motivates the forthcoming factorization. For p a prime and n a nonzero integer, define $\nu_p(n) = d$ if $p^d||n$. Then we set $c_2 = c_z c_f c_g$ where

$$c_z = \prod_{\substack{p^\beta || c_2 \\ 1 \le \nu_p(c_2') < \nu_p(c_2'')}} p^\beta, \qquad c_f = \prod_{\substack{p^\beta || c_2 \\ \nu_p(c_2'') \le \nu_p(c_2')}} p^\beta, \qquad c_g = \prod_{\substack{p^\beta || c_2 \\ \nu_p(c_2')=0 \\ \nu_p(c_2'') \ge 1}} p^\beta.$$

According to this factorization, we also write $c_2' = c_z' c_f' c_g'$ and $c_2'' = c_z'' c_f'' c_g''$ where $c_*' = (c_*, c_2')$ and $c_*'' = (c_*, c_2'')$ with $* = z, f, g$. Note from the definitions that c_z, c_f, c_g are all pairwise relatively prime, and also that $c_f''|c_f'$, and $c_g' = 1$.

Using the Chinese remainder theorem, we factor $G_{A,B}$ as

$$\begin{aligned} G_{A,B}(a_1, a_2, a_3; c_2) = \frac{(A, c_2)}{c_2} & \sum_{\substack{u_e \pmod{c_z} \\ u_e \equiv -\overline{B'}\widetilde{a_3} \pmod{c_z'}}}^{*} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2} u_e (c_f c_g)\overline{c_f c_g}}{c_z'' c_f'' c_g''}\right) \\ & \times \sum_{\substack{u_f \pmod{c_f} \\ u_f \equiv -\overline{B'}\widetilde{a_3} \pmod{c_f'}}}^{*} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2} u_f (c_z c_g)\overline{c_z c_g}}{c_z'' c_f'' c_g''}\right) \\ & \times \sum_{\substack{u_g \pmod{c_g} \\ u_g \equiv -\overline{B'}\widetilde{a_3} \pmod{c_g'}}}^{*} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2} u_g (c_z c_f)\overline{c_z c_f}}{c_z'' c_f'' c_g''}\right). \end{aligned}$$

Let us examine each of these three sums in turn. We begin by writing the sum over u_e more suggestively as follows:

$$\sum_{\substack{u \pmod{c_z} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_z'}}}^{*} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2} u (\frac{c_f c_g}{c_f'' c_g''})\overline{c_f c_g}}{c_z''}\right).$$

For each prime p dividing c_z, we have $p|c_z'$, and $p|\frac{c_z''}{c_z'}$. Therefore, we may write $u = -\overline{B'}\widetilde{a_3} + c_z' t$, where t runs over *all* residue classes modulo c_z''/c_z'. But then the sum over t vanishes, because the factor in the numerator is relatively prime to the denominator. Thus we obtain that G vanishes unless $c_z = 1$. We henceforth make this assumption in the next computations of the c_f and c_g moduli sums.

For the sum over $u_f \pmod{c_f}$, since $c_f''|c_f'$, the congruence uniquely determines u_f modulo c_f'', so we get, using $(c_g/c_g'')\overline{c_g} = \overline{c_g''}$:

$$\sum_{\substack{u \pmod{c_f} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_f'}}}^{*} e\left(\frac{-\overline{A'}\widetilde{a_1}\widetilde{a_2} u c_g \overline{c_g}}{c_f'' c_g''}\right) = e\left(\frac{\overline{A'B'}\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}\overline{c_g''}}{c_f''}\right) \sum_{\substack{u \pmod{c_f} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c_f'}}}^{*} 1.$$

We also have

$$\sum_{\substack{u \pmod{c_f} \\ u \equiv -\overline{B'}\widetilde{a_3} \pmod{c'_f}}}^{*} 1 = \frac{\varphi(c_f)}{\varphi(c'_f)},$$

as can be checked as follows. Firstly, we see that both sides of the purported identity are multiplicative, so it suffices to check this on prime powers. If $c'_f = 1$, then the identity follows easily. If $c'_f = p^{\beta'}$, and $c_f = p^{\beta}$, and $\beta' \geq 1$, then the left hand side is $p^{\beta-\beta'} = \frac{\varphi(p^{\beta})}{\varphi(p^{\beta'})}$, as desired.

Finally, we examine the sum modulo c_g. We have $c'_g = 1$ (directly from the definition, as remarked earlier), so the congruence condition is vacuous. The sum then simplifies as $R_{c_g}(\frac{c_g}{c''_g})$, using that c_f and everything else in the numerator in the exponential is relatively prime to c_g.

Putting all these calculations together, we have shown that if $c_2 = c_f c_g$, $(B, c_2) = (a_3, c_2)$, and $(A, c_2) \mid (a_1, a_2)$, then

$$G_{A,B}(a_1, a_2, a_3; c_2) = \frac{(A, c_2)}{c_2} R_{c_g}\left(\frac{c_g}{c''_g}\right) \frac{\varphi(c_f)}{\varphi(c'_f)} e\left(\frac{\overline{A'B'}\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}\overline{c''_g}}{c''_f}\right), \tag{111}$$

and otherwise $G_{A,B}$ vanishes. So far the assumption $a_1 a_2 a_3 \neq 0$ was not used.

To estimate this expression for G, we have $c'_f = (c'_2, c_f) = (\frac{c_2}{(B,c_2)}, \frac{c_2}{c_g})$ and $c_g \mid (B, c_2)$, so in fact $c'_f = \frac{c_2}{(B,c_2)} = c'_2$. Then

$$\left| R_{c_g}\left(\frac{c_g}{c''_g}\right) \right| \frac{\varphi(c_f)}{\varphi(c'_f)} \leq \varphi(c_g) \frac{\varphi(c_f)}{\varphi(c'_f)} = \frac{\varphi(c_2)}{\varphi(c_2/(c_2, B))} \leq (B, c_2). \tag{112}$$

Our goal now is to use Dirichlet characters to decompose

$$e_{c_2}(-a_1 a_2 a_3) G_{A,B}(a_1, a_2, a_3; c_2).$$

Switching to the new notation used in (111), we have

$$e\left(\frac{-a_1 a_2 a_3}{ABc_2}\right) = e\left(\frac{-\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}}{A'B'c''_f c''_g}\right),$$

and by reciprocity, we have

$$c_2 e\left(\frac{-a_1 a_2 a_3}{ABc_2}\right) G_{A,B}(a_1, a_2, a_3; c_2) = (A, c_2) R_{c_g}\left(\frac{c_g}{c''_g}\right) \frac{\varphi(c_f)}{\varphi(c'_f)} e\left(\frac{-\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}\overline{c''_f}}{c''_g A'B'}\right).$$

Let

$$g_3 = (\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}, c''_g A'B').$$

Then we have

$$c_2 e\left(\frac{-a_1a_2a_3}{ABc_2}\right) G_{A,B}(a_1,a_2,a_3;c_2)$$
$$= (A,c_2) R_{c_g}\left(\frac{c_g}{c_g''}\right)\frac{\varphi(c_f)}{\varphi(c_f')}\frac{1}{\varphi\left(\frac{c_g''A'B'}{g_3}\right)}\sum_{\eta \left(\mathrm{mod}\ \frac{c_g''A'B'}{g_3}\right)} \tau(\overline{\eta})\eta(-\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}\overline{c_f''}/g_3).$$

Finally, we argue that this expression is of the desired form for Lemma 6. Recall we write $a_i = u_i v_i$ where $(u_i, AB) = 1$ and $v_i | (AB)^\infty$. As originally written, the g_i depend on the a_i, but in fact they only depend on the v_i since the g_i are divisors of c_2, and $c_2 | (AB)^\infty$. By writing the dependence of the g_i on the v_i explicitly as summation conditions, we see the presence of the first sum in (101). A careful scrutiny of the changes of variables throughout the proof shows that the variables $c_g, c_g'', c_f, c_f', A', B', g_3$ are functions of c_2, the v_i and g_1, g_2, and are independent of the u_i. We may also extract from $\widetilde{a_1}\widetilde{a_2}\widetilde{a_3}$ the factor $u_1u_2u_3$. We obtain the bound on γ_η by (112), and using the standard bound on the Gauss sum. We note that

$$\frac{c_g''A'B'}{g_3} | c_2''A'B', \quad \text{and} \quad c_2''A'B' | \frac{c_2AB}{(A,c_2)^2(B,c_2)},$$

which gives the divisibility condition on D.

The only remaining statement to prove is (103). In this case, the Gauss sum is bounded by 1, and by (112) we have

$$(A,c_2)\left|R_{c_g}\left(\frac{c_g}{c_g''}\right)\right|\frac{\varphi(c_f)}{\varphi(c_f')}\frac{1}{\varphi\left(\frac{c_g''A'B'}{g_3}\right)} \ll (qr)^\varepsilon (A,c_2)(B,c_2)\frac{g_3}{c_g''A'B'}.$$

By tracing back the definitions, we see that

$$g_3 = \left(\frac{v_1v_2v_3}{g_1g_2(A,c_2)^2(B,c_2)}, c_g''A'B'\right) \le c_g''\left(\frac{v_1v_2v_3}{(A,c_2)^2(B,c_2)}, A'B'\right),$$

which implies the bound (103). □

Proof (of Lemma 7) Now we evaluate $G_{1,1}(a_1,a_2,a_3;c_e)$. To do so we break into cases. First assume that $q_e \nmid c_e$. Under the condition $q_e \nmid c_e$ there are only finitely many possibilities for q_e, c_e, a_1, a_2, a_3. A brute force computation with Sage [42] then shows that if $c_e = 1, 2$ then

$$G_{1,1}(a_1,a_2,a_3;c_e) = \frac{1}{q_e^2}\chi_{q_e}(-1)\tau(\chi_{q_e})\chi_{q_e}(a_1a_2a_3), \tag{113}$$

and if $q_e = 8$ and $c_e = 4$ then

$$G_{1,1}(a_1, a_2, a_3; c_e) = \frac{1}{32}\chi_{q_e}(-1)\tau(\chi_{q_e}\chi_4)\chi_{q_e}\chi_4(a_1a_2a_3). \tag{114}$$

Now we assume that $q_e \mid c_e$. We write $c_e = q_e s_e$ where s_e is a power of 2. Following the same steps as [14, Section 10] we have that if $q_e \mid c_e$ then

$$G_{1,1}(a_1, a_2, a_3; c_e) = \frac{\delta_{(a_3,s_e)=1}\chi_{q_e}(-1)}{q_e^2 s_e} e\left(\frac{a_1a_2a_3}{c_e}\right) H_{s_e}(a_1, a_2, a_3, q_e), \tag{115}$$

where H_s is defined in [14, (10.2)]. Assume now both $q_e \mid c_e$ and $q_e \mid s_e$ (so that in fact $q_e^2 \mid c_e$). Following the proof of Lemma 10.1 of [14] we find in this case that

$$H_{s_e}(a_1, a_2, a_3, q_e) = \chi_{q_e}(a_1a_2a_3)^2\tau(\chi_{q_e})^2. \tag{116}$$

Having dealt with this case, we may now assume that $q_e \mid c_e$ and $q_e \nmid s_e$. Now there are only finitely many choices for q_e, c_e, a_1, a_2, a_3 which permits us to conclude the following lemma by another Sage computation.

Lemma 8 *Suppose that $q_e \mid c_e$ and let $s_e = c_e/q_e$. If $s_e = 1$ we have*

$$H_{s_e}(a_1, a_2, a_3, q_e) = \begin{cases} \frac{1}{2}R_{q_e}(a_1)R_{q_e}(a_2)R_{q_e}(a_3) & \text{if } q_e = 4 \\ \frac{1}{4}R_{q_e}(a_1)R_{q_e}(a_2)R_{q_e}(a_3) & \text{if } q_e = 8 \text{ and } 4|a_1, a_2, a_3, \\ 16i\,\chi_4\left(\frac{a_1a_2a_3}{8}\right) & \text{if } q_e = 8 \text{ and } 2||a_1, a_2, a_3, \\ 0 & \text{otherwise.} \end{cases} \tag{117}$$

If $2 \mid s_e$ then

$$H_s(a_1, a_2, a_3, q_e) = \begin{cases} \chi_{q_e}(a_1a_2a_3)^2\tau(\chi_{q_e})^2 & \text{if } q_e \mid s_e \\ -\chi_{q_e}(a_1a_2a_3)^2\tau(\chi_{q_e})^2 & \text{if } 2s_e = q_e \\ i\tau(\chi_{q_e})^2\chi_4(a_1a_2a_3) & \text{if } s_e = 2 \text{ and } q_e = 8. \end{cases} \tag{118}$$

If $q_e \mid c_e$ then the additive character on the left hand side of Lemma 7 cancels identically with the additive character appearing in (115). On the other hand, if $q_e \nmid c_e$ then the additive character $e_{[c_e,q_e]^3/c_e^2}(-a_1a_2a_3) = e_{q_e^3/c_e^2}(-a_1a_2a_3)$ must be expressed in terms of multiplicative characters. Recall, if $a_1a_2a_3 \neq 0$ we factor $a_i = e_i f_i$ with e_i a power of 2 and f_i odd. We have

$$e_{q_e^3/c_e^2}(-a_1a_2a_3) = \frac{1}{\varphi(q_e^3/c_e^2(q_e^3/c_e^2, a_1a_2a_3))} \sum_{\theta \left(\mathrm{mod}\ \frac{q_e^3}{c_e^2(q_e^3/c_e^2, a_1a_2a_3)}\right)} \tau(\overline{\theta})\theta\left(-\frac{a_1a_2a_3}{(q_e^3/c_e^2, a_1a_2a_3)}\right), \tag{119}$$

where by convention we take $\theta \pmod 1$ to be identically 1.

Having computed $G_{1,1}(a_1, a_2, a_3; c_e)$ we now argue that the resulting expressions are of the desired form for Lemma 7. In similar fashion to the proof of Lemma 6, note that for each fixed q_e, c_e we have that $c_e q_e G'(a_1, a_2, a_3; c_e)$ is of the form (104) by inspecting (119), (113), (114), (117) and (118). Now allowing the function $\mathfrak{g}$ to depend on q_e, c_e and seeing that $\frac{q_e^3}{c_e^2(q_e^3/c_e^2, a_1a_2a_3)} \mid 64$ in all cases, we conclude the statement of Lemma 7. □

Proof (of Proposition 3) Let

$$G'_{A,B}(m_1, m_2, m_3; c) = G_{A,B}(m_1, m_2, m_3; c) e\left(\frac{-m_1m_2m_3}{AB[c, q]^3/c^2}\right).$$

Then by (97) and (98) we have

$$\begin{aligned} cqG'_{A,B}(m_1, m_2, m_3; c) = \chi_q(AB)c_oq_oG'_{1,1}\left(m_1, m_2, \overline{ABc_2[c_e, q_e]}^3 c_e^2 m_3; c_o\right) \\ \times c_2 G'_{A,B}\left(m_1, m_2, \overline{c_o[c_e, q_e]}^3 c_e^2 m_3; c_2\right) \\ \times c_e q_e G'_{1,1}\left(m_1, m_2, \overline{ABc_2c_o} m_3; c_e\right). \end{aligned}$$

Now we factor $m_i = m_i' m_i''$, where $(m_i', AB) = 1$ and $m_i'' \mid (AB)^\infty$, and then further factor $m_i' = m_i^o m_i^e$ where m_i^o is odd and m_i^e is a power of 2. By Lemmas 5, 6, and 7, we have

$$\begin{aligned} &cqG'_{A,B}(m_1, m_2, m_3; c) \\ &= \delta\left((A, c_2) \mid \left(m_1'', m_2''\right)\right)\delta\left((B, c_2) \mid m_3''\right) \sum_{\substack{g_1g_2 \mid \frac{c_2}{(A,c_2)} \\ g_1 = \left(\frac{m_1''}{(A,c_2)}, \frac{c_2}{(A,c_2)}\right) \\ g_2 = \left(\frac{m_2''}{(A,c_2)}, \frac{c_2}{g_1(A,c_2)}\right)}} \sum_{\substack{D_1D_2hk = q_o \\ h = (q_o, s_o) \\ k = (m_1^o m_2^o m_3^o, q) \\ (h, m_1^o m_2^o) = 1 \\ (m_3^o, c_o/q_o) = 1}} \frac{1}{\varphi(D_2)} \\ &\times \sum_{\psi \,(\mathrm{mod}\ D_2)} \sum_{\Delta \mid 64} \frac{1}{\varphi(\Delta)} \sum_{\chi \,(\mathrm{mod}\ \Delta)} \sum_{D \mid \frac{c_2AB}{(c_2,A)^2(c_2,B)}} \frac{1}{\varphi(D)} \sum_{\eta \,(\mathrm{mod}\ D)} G_* \\ &\times (\psi\chi\eta)\left(m_1^o m_2^o m_3^o \overline{s_o}\right) R_k\left(m_1^o\right) R_k\left(m_2^o\right) R_k\left(m_3^o\right), \end{aligned} \tag{120}$$

where G_* is the product of the g, γ, and $\mathfrak{g}$ arising in Lemmas 5, 6, and 7 along with various miscellaneous factors of unit size, such as $\chi_q(AB)$. The exact form of G_* is not important. Rather, all that matters is a bound on its absolute value, and the fact that it does *not* depend on m_1^o, m_2^o, m_3^o, s_o, q_o, c_o. Specifically, we have it is of size

$$G_* \ll D^{1/2}(A, c_2)(B, c_2)D_2^{3/2+\varepsilon} \tag{121}$$

and if ψ, χ, and η are all the principal character then with $A = (A, c_2)A'$ and $B = (B, c_2)B'$

$$\frac{G_*}{D} \ll (qr)^{\varepsilon}(A,c_2)(B,c_2)\frac{\left(\frac{m_1''m_2''m_3''}{(A,c_2)^2(B,c_2)}, A'B'\right)}{A'B'}. \tag{122}$$

Now we are ready to sum $G'_{A,B}$ over the m_i and c. We must break into two cases to handle the condition $(c, q_e) = \delta$.

First suppose that q is odd i.e., $q_e = 1$. Then since $(c, q_e) = 1$ we have that the sum over c is empty unless $\delta = 1$, and if $\delta = 1$ the the condition $(c, q_e) = \delta$ is true for all c. We factor $R = R_eR_o$ where R_o is odd and R_e is a power of 2 and write $c_o = t_oqR_o$ and $c_e = t_eR_e$. Then for any function f for which the sums converge absolutely we have

$$\sum_{\substack{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)\\(c,q_e)=\delta}} f(c) = \delta(\delta=1)\sum_{c_2|(AB)^\infty}\sum_{\substack{t_e|2^\infty\\(t_e,AB)=1}}\sum_{(t_o,2AB)=1} f(c_2t_oqR_ot_eR_e).$$

Now we suppose that $q_e = 4$ or 8. Then R must be odd as $(R, q) = 1$. Recall that $(AB, qR) = 1$ so that $(c_2, R\tilde{q}) = 1$, and also that $\tilde{q}/q_o = 2$. Then we write $c_o = t_oq_oR$ and $c_e = 2t_e$. Then we have

$$\sum_{\substack{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)\\(c,q_e)=\delta}} f(c) = \delta(\delta \mid q_e)\sum_{c_2|(AB)^\infty}\sum_{\substack{t_e|2^\infty\\(t_e,AB)=1\\ \delta|2t_e\\ \left(\frac{2t_e}{\delta},\frac{q_e}{\delta}\right)=1}}\sum_{(t_o,2AB)=1} f(2c_2t_oq_oRt_e),$$

for any f for which the sums converge absolutely. We treat only this last case for the remainder of the section, as the other cases are strictly simpler. Applying this decomposition to $G'_{A,B}$ we find if $\delta \mid q_e$ that

$$\begin{aligned}
Z_{\delta,R,q} &= \Bigg[\sum_{c_2|(AB)^\infty}\sum_{\substack{t_e|2^\infty\\(t_e,AB)=1\\ \delta|2t_e\\ \left(\frac{2t_e}{\delta},\frac{q_e}{\delta}\right)=1}}\sum_{m_1'',m_2'',m_3''|(AB)^\infty}\sum_{\substack{m_1^e,m_2^e,m_3^e|2^\infty\\(m_1^em_2^em_3^e,AB)=1}}\\
&\qquad\times\sum_{(t_o,2AB)=1}\sum_{(m_1^om_2^om_3^o,2AB)=1}\frac{cqG'_{A,B}(m_1,m_2,m_3;c)}{m_1^{s_1}m_2^{s_2}m_3^{s_3}(c_2t_et_o)^{s_4}}\Bigg]\\
&= \Bigg[\sum_{\substack{c_2,t_e,m_1'',m_2'',m_3'',m_1^e,m_2^e,m_3^e\\(A,c_2)|(m_1'',m_2''),\,(B,c_2)|m_3'',(\dots)}}\sum_{\substack{g_1g_2|\frac{c_2}{(A,c_2)}\\ g_1=\left(\frac{m_1''}{(A,c_2)},\frac{c_2}{(A,c_2)}\right)\\ g_2=\left(\frac{m_2''}{(A,c_2)},\frac{c_2}{g_1(A,c_2)}\right)}}\sum_{\Delta|64}\frac{1}{\varphi(\Delta)}\sum_{\chi\,(\mathrm{mod}\ \Delta)}\\
&\qquad\times\sum_{D_1D_2hk=q_o}\frac{1}{\varphi(D_2)}\sum_{\psi\,(\mathrm{mod}\ D_2)}\sum_{D|\frac{c_2AB}{(c_2,A)^2(c_2,B)}}\frac{1}{\varphi(D)}\sum_{\eta\,(\mathrm{mod}\ D)}G'_*Y\Bigg],
\end{aligned} \tag{123}$$

where G'_* satisfies (121) and (122), the conditions $(\dots)$ in the first (large) summand following the second equals sign are the same conditions as on the first line of (123), and

$$Y = \sum_{\substack{(m_1^o m_2^o m_3^o t_o, 2AB)=1 \\ h=(q_o, R_o t_o) \\ k=(m_1^o m_2^o m_3^o, q_o) \\ (h, m_1^o m_2^o)=1 \\ (m_3^o, R_o t_o)=1}} \frac{(\psi\chi\eta)(m_1^o m_2^o m_3^o \overline{t_o}) R_k(m_1^o) R_k(m_2^o) R_k(m_3^o)}{(m_1^o)^{s_1} (m_2^o)^{s_2} (m_3^o)^{s_3} t_o^{s_4}}.$$

Our next goal is to obtain meromorphic continuation of Y inside the critical strip, and a bound on Y both slightly to the right of the critical lines, and slightly to the right of the edge of absolute convergence. First we note the following formal combinatorial identity:

$$\sum_{(n_1 n_2 n_3, q_o)=k} f(n_1, n_2, n_3) = \sum_{k_1 k_2 k_3 = k} \sum_{\left(n_1, \frac{q_o}{k_1}\right)=1} \sum_{\left(n_2, \frac{q_o}{k_1 k_2}\right)=1} \sum_{\left(n_3, \frac{q_o}{k_1 k_2 k_3}\right)=1} f(k_1 n_1, k_2 n_2, k_3 n_3).$$

With this, we have (with some minor simplifications arising from $(q_o, AB) = 1$ which means for instance that $(k, AB) = 1$)

$$Y = \sum_{k_1 k_2 k_3 = k} \frac{(\psi\chi\eta)(k_1 k_2 k_3)}{k_1^{s_1} k_2^{s_2} k_3^{s_3}} \sum_{\substack{(t_o, k_3 AB)=1 \\ (t_o, q_o)=h}} \times \sum_{\substack{\left(m_1^o, hAB\frac{q_o}{k_1}\right)=1 \\ \left(m_2^o, hAB\frac{q_o}{k_1 k_2}\right)=1 \\ \left(m_3^o, R_o t_o AB\frac{q_o}{k}\right)=1}} \frac{(\psi\chi\eta)(m_1^o m_2^o m_3^o \overline{t_o})}{(m_1^o)^{s_1} (m_2^o)^{s_2} (m_3^o)^{s_3} t_o^{s_4}} R_k(k_1 m_1^o) R_k(k_2 m_2^o) R_k(k_3 m_3^o). \tag{124}$$

The condition $(k_1 k_2, h) = 1$ is automatic, because $k_1 k_2 k_3 = k$, $(s_o, q_o) = h$, $hk \mid q_o$, and q_o is square-free, so $(h, k) = 1$.

Now let $a, b \in \mathbb{N}$ and suppose that $a'' \mid a' \mid a$, with a' square-free, and χ is a Dirichlet character mod a. Then for $\mathrm{Re}(s) > 1$ we have

$$\sum_{(n,b)=1} \frac{R_{a'}(a''n)\chi(n)}{n^s} = \mu(a'/a'')\phi(a'')L_b(s, \chi), \tag{125}$$

where $L_b(s, \chi)$ is the Dirichlet L-function with Euler factors at primes dividing b omitted. To see this, observe that if a' is square-free then $\mu(a')R_{a'}(n)$ is a multiplicative function of n, and that the summand on the left side of (125) vanishes whenever

$(n, a'') \neq 1$ because $a'' \mid a$. One can then factor out a Ramanujan sum from the left hand side and use the fact that $\mu(a')R_{a'}(n) = 1$ if $(n, a') = 1$.

From this, we easily get the meromorphic continuation of Y to, say $\mathrm{Re}(s_i) > 1/2$, $i = 1, 2, 3, 4$. Moreover, Y is analytic except for possible poles at $s_i = 1$ in case $\eta\psi\chi$ is the principal character (which then implies all of χ, η, and ψ are principal, since their respective moduli are coprime). Assuming $\mathrm{Re}(s_i) = \sigma > 1/2$ for all $i = 1, 2, 3, 4$, and $\sigma \neq 1$, we have

$$|Y| \ll_\sigma k^{1-\sigma} h^{-\sigma} (qr)^\varepsilon |L(s_1, \psi\chi\eta)L(s_2, \psi\chi\eta)L(s_3, \psi\chi\eta)L(s_4, \overline{\psi\chi\eta})|. \quad (126)$$

Now let $Z_{\delta,R,q} = Z_0 + Z'$ where Z_0 corresponds to the terms with $\eta\chi\psi$ principal, and Z' corresponds to the terms with $\eta\chi\psi$ nonprincipal.

Taking $\sigma = 1 + \varepsilon$, we bound Z_0 as follows:

$$|Z_0| \ll (qr)^\varepsilon \sum_{\substack{m_1'', m_2'', m_3'', c_2 \mid (AB)^\infty \\ (A,c_2)\mid (m_1'', m_2'') \\ (B,c_2)\mid m_3''}} \frac{1}{(m_1'' m_2'' m_3'' c_2)^\sigma} \sum_{D_1 D_2 hk = q_o} \sum_{\substack{g_1 g_2 \mid \frac{c_2}{(A,c_2)} \\ g_1 = \left(\frac{m_1''}{(A,c_2)}, \frac{c_2}{(A,c_2)}\right) \\ g_2 = \left(\frac{m_2''}{(A,c_2)}, \frac{c_2}{g_1(A,c_2)}\right)}}$$
$$\times \frac{1}{\varphi(D_2)} \sum_{D \mid \frac{c_2 AB}{(c_2,A)^2(c_2,B)}} (A, c_2)(B, c_2) \frac{\left(\frac{m_1'' m_2'' m_3''}{(A,c_2)^2(B,c_2)}, A'B'\right)}{A'B'} k^{1-\sigma} h^{-\sigma}.$$

Next we change variables $m_i'' = n_i(A, c_2)$ for $i = 1, 2$ and $m_3'' = (B, c_2)n_3$. We have

$$\sum_{n \mid (AB)^\infty} \frac{(n, Q)}{n} = \prod_{p \mid AB} \sum_{j=0}^{\infty} \frac{(p^j, Q)}{p^j} \ll (ABQ)^\varepsilon. \quad (127)$$

Using this successively on n_1, n_2, n_3, and trivially summing over g_1, g_2, D, we obtain

$$|Z_0| \ll (qr)^\varepsilon \sum_{c_2 \mid (AB)^\infty} \frac{1}{c_2^\sigma} \sum_{D_1 D_2 hk = q_o} \frac{1}{\varphi(D_2)} \frac{(A, c_2)(B, c_2)}{(A, c_2)^2 (B, c_2)} \frac{1}{A'B'} k^{1-\sigma} h^{-\sigma}.$$

We use the estimate (127) again on the sum over c_2 to get

$$|Z_0| \ll \frac{(qr)^\varepsilon}{AB} \sum_{D_1 D_2 hk = q_o} k^{1-\sigma} h^{-\sigma} D_2^{-1} \ll \frac{(qr)^\varepsilon}{AB}.$$

This proves the bound (94), as desired.

Next we turn to Z'. For this, we use the large sieve inequality to give a bound on the 4th moment of Dirichlet L-functions. Following e.g. [40, Lemma 8] we find that for $\sigma = 1/2 + \varepsilon$, we have

$$\frac{1}{\phi(Q)} \sum_{\chi \,(\mathrm{mod}\ Q)} |L(s,\chi)|^4 \ll Q^{\varepsilon}(1+|s|)^{1+\varepsilon}. \tag{128}$$

Using Hölder's inequality, we have for $\sigma_i = 1/2+\varepsilon$ for $i = 1,2,3,4$, that

$$|Z'| \ll (qr)^{\varepsilon} \sum_{\substack{m_1'',m_2'',m_3'',c_2 | (AB)^{\infty} \\ (A,c_2)|(m_1'',m_2'') \\ (B,c_2)|m_3''}} \frac{1}{(m_1''m_2''m_3''c_2)^{1/2+\varepsilon}} \sum_{D_1D_2hk=q} \sum_{\substack{g_1g_2|\frac{c_2}{(A,c_2)} \\ g_1=\left(\frac{m_1''}{(A,c_2)},\frac{c_2}{(A,c_2)}\right) \\ g_2=\left(\frac{m_2''}{(A,c_2)},\frac{c_2}{g_1(A,c_2)}\right)}}$$

$$\times \sum_{D|\frac{c_2AB}{(c_2,A)^2(c_2,B)}} (A,c_2)(B,c_2) D^{1/2} D_2^{3/2} k^{1-\sigma} h^{-\sigma} \prod_{j=1}^{4} (1+|s_j|)^{1/4+\varepsilon}.$$

Using similar methods to estimate the sums over the m_i'' and c_2 as in the bound on Z_0, we obtain

$$|Z'| \ll (qr)^{\varepsilon} q^{3/2+\varepsilon} (AB)^{1/2} \prod_{j=1}^{4} (1+|s_j|)^{1/4+\varepsilon}.$$

Finally, we show (96). The proof is essentially the same as before, except we use a hybrid large sieve in place of (128), as in [20], namely

$$\int_{|t|\le T} \sum_{\chi \,(\mathrm{mod}\ q)} |L(1/2+it,\chi)|^4 dt \ll (qT)^{1+\varepsilon}.$$

□

We conclude this section by studying $G_{A,B}(m_1,m_2,m_3;c)$ when some $m_i = 0$. The formulas greatly simplify.

Lemma 9 *Suppose some $a_j = 0$. If $a_i \neq 0$ write $a_i = a_i''a_i'$ where $a_i'' \mid (2AB)^{\infty}$ and $(a_i', 2AB) = 1$, and if $a_i = 0$ write $a_i'' = a_i' = 0$. Then*

$$cq\phi(q) G_{A,B}(a_1,a_2,a_3;c) = g\left(a_1'',a_2'',a_3'',A,B,c,q\right) R_{q_o}\left(a_1'\right) R_{q_o}\left(a_2'\right) R_{q_o}\left(a_3'\right),$$

where g is a function satisfying the bound

$$\left|g\left(a_1'',a_2'',a_3'',A,B,c,q\right)\right| \le 64(A,c)(B,c).$$

Proof We have according to (97) and (98) that

$$G_{A,B}(a_1,a_2,a_3;c) = \chi_q(AB) G_{1,1}\left(a_1,a_2,\overline{ABc_2[c_e,q_e]}^3 c_e^2 a_3;c_o\right)$$
$$\times G_{1,1}\left(a_1,a_2,\overline{ABc_2c_o}a_3;c_e\right) G_{A,B}\left(a_1,a_2,\overline{[c_1,q]}^3 c_1^2 a_3;c_2\right).$$

We treat each of these in turn. Let us begin with $G_{1,1}(a_1, a_2, \overline{ABc_2[c_e, q_e]}^3 c_e^2 a_3; c_o)$. The evaluation (105) of $G_{1,1}$ did not require $a_1a_2a_3 \neq 0$. Inspecting (105), and noting $h = 1$, $k = q_o$ under the assumption $a_1a_2a_3 = 0$, we have

$$G_{1,1}\left(a_1, a_2, \overline{ABc_2[c_e, q_e]}^3 c_e^2 a_3; c_o\right) = \begin{cases} \frac{\chi_{q_o}(-1)}{c_o q_o \phi(q_o)} R_{q_o}(a_1) R_{q_o}(a_2) R_{q_o}(a_3) & \text{if } \left(\frac{c_o}{q_o}, a_3\right) = 1 \text{ and } (q_o, \frac{c_o}{q_o}, a_1a_2) = 1 \\ 0 & \text{else.} \end{cases}$$

Next consider $G_{1,1}(a_1, a_2, \overline{ABc_2c_o}a_3; c_e)$. An inspection of the proof of Lemma 7 shows that when $a_1a_2a_3 = 0$ that

$$G_{1,1}\left(a_1, a_2, \overline{ABc_2c_o}a_3; c_e\right) = \begin{cases} \frac{1}{q_e c_e \phi(q_e)} \chi_{q_e}(-1) R_{q_e}(a_1) R_{q_e}(a_2) R_{q_e}(a_3) & \text{if } q_e = c_e \\ \frac{1}{q_e c_e \phi(q_e)} \delta_{(a_3, c_e)=1} & \text{if } q_e = 1 \\ 0 & \text{else,} \end{cases}$$

where note that in the case $q_e = c_e = 8$, $a_1a_2a_3 = 0$, the third case in (117) may be discarded, and in the second case of (117), the condition $4|a_1, a_2, a_3$ may be dropped since the Ramanujan sum vanishes otherwise. This function only depends on q_e,c_e, and the 2-part of a_1, a_2, a_3 and is bounded above by $64/q_e c_e \phi(q_e)$, since each Ramanujan sum is bounded by 4 in absolute value.

Lastly, consider $G_{A,B}(a_1, a_2, \overline{[c_1, q]}^3 c_1^2 a_3; c_2)$. As mentioned in the proof of Lemma 6, (111) is valid without the assumption that $a_1a_2a_3 \neq 0$. We have that $G_{A,B}(a_1, a_2, \overline{[c_1, q]}^3 c_1^2 a_3, c_2)$ only depends on a_1'', a_2'', a_3'', A, B, and c_2. In particular, by (112) we have

$$|G_{A,B}(a_1, a_2, \overline{[c_1, q]}^3 c_1^2 a_3; c_2)| \leq \frac{(A, c_2)(B, c_2)}{c_2}.$$

Therefore

$$G_{A,B}(a_1, a_2, a_3; c) = g(a_1'', a_2'', a_3'', A, B, c, q) \frac{1}{cq\phi(q)} R_q(a_1') R_q(a_2') R_q(a_3'),$$

and $|g(a_1'', a_2'', a_3'', A, B, c, q)| \leq (A, c_2)(B, c_2)$. □

10 Weight functions

10.1 Inert functions

We begin this section by quoting a definition of [28].

Let $\mathcal{F}$ be an index set and $X = X_T : \mathcal{F} \to \mathbb{R}_{\geq 1}$ be a function of $T \in \mathcal{F}$.

Definition 1 A family $\{w_T\}_{T\in\mathcal{F}}$ of smooth functions supported on a product of dyadic intervals in $\mathbb{R}^d_{>0}$ is called X-inert if for each $j=(j_1,\dots,j_d)\in\mathbb{Z}^d_{\geq 0}$ we have

$$C(j_1,\dots,j_d) := \sup_{T\in\mathcal{F}}\sup_{(x_1,\dots,x_d)\in\mathbb{R}^d_{>0}} X_T^{-j_1-\cdots-j_d}\left|x_1^{j_1}\cdots x_d^{j_d} w_T^{(j_1,\dots,j_d)}(x_1,\dots,x_d)\right| < \infty. \tag{129}$$

We often abbreviate the sequence of constants $C(j_1,\dots,j_d)$ associated to a family of inert functions by $C_{\mathcal{F}}$. We will most often use the definition of X-inert with $X = X_T = 1$, although occasionally (e.g., in the proof of Lemma 12) X will be slightly larger than 1, with $X \ll (qr)^{\varepsilon}$.

The purpose of this definition is to encode natural conditions on a weight function that lets us separate variables efficiently. For instance, if w_T satisfies (129), then by Mellin inversion,

$$w_T(x_1,\dots,x_d) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\widetilde{w_T}(it_1,\dots it_d)x_1^{-it_1}\dots x_d^{-it_d}dt_1\dots dt_d, \tag{130}$$

where

$$\widetilde{w_T}(s_1,\dots,s_d) = \int_{(0,\infty)^d} w_T(x_1,\dots,x_d)x_1^{s_1}\dots x_d^{s_d}\frac{dx_1}{x_1}\dots\frac{dx_d}{x_d}.$$

Integrating by parts shows for any choices of $j_1,\dots,j_d = 0,1,\dots,$ we have

$$\widetilde{w_T}(s_1,\dots,s_d) = \left(\prod_{a=1}^{d}\prod_{b=0}^{j_a-1}\frac{-1}{s_a+b}\right)\int_{(0,\infty)^d} w_T^{(j_1,\dots,j_d)}(x_1,\dots,x_d)x_1^{s_1+j_1}\dots x_d^{s_d+j_d}\frac{dx_1}{x_1}\dots\frac{dx_d}{x_d}.$$

Therefore, by (129), we have

$$|\widetilde{w_T}(it_1,\dots,it_d)| \leq \left(\frac{X_T}{|t_1|}\right)^{j_1}\dots\left(\frac{X_T}{|t_d|}\right)^{j_d} C(j_1,\dots,j_d)(\log 2)^d.$$

If $|t_i|\geq X_T$, then we take j_i as unspecified (arbitrarily large), while if $|t_i|<X_T$, we choose $j_i=0$. In this way, we obtain

$$|\widetilde{w_T}(it_1,\dots,it_d)| \leq \left(1+\frac{|t_1|}{X_T}\right)^{-j_1}\dots\left(1+\frac{|t_d|}{X_T}\right)^{-j_d} C'(j_1,\dots,j_d), \tag{131}$$

where C' is some other sequence depending only on C. Our interpretation of this estimate combined with (130) is that w_T can have its variables separated "at cost" X_T^d, meaning that each integral has essential length $\ll X_T$.

10.2 Integration by parts

Often an integral can be shown to be small by repeated integration by parts. For this, we quote [10, Lemma 8.1] with some slight changes of notation and terminology.

Lemma 10 [10] *Suppose that $w = w_T(t)$ is a family of X-inert functions, with compact support on $[Z, 2Z]$, so that $w^{(j)}(t) \ll (Z/X)^{-j}$. Also suppose that ϕ is smooth and satisfies $\phi^{(j)}(t) \ll \frac{Y}{Z^j}$ for some $Y/X^2 \geq R \geq 1$ and all t in the support of w. Let*

$$I = \int_{-\infty}^{\infty} w(t) e^{i\phi(t)} dt.$$

If $|\phi'(t)| \gg \frac{Y}{Z}$ for all t in the support of w, then $I \ll_A ZR^{-A}$ for A arbitrarily large.

10.3 Stationary phase

Now we quote the main theorem from [28] which extends [10, Proposition 8.2].

Theorem 6 [28] *Suppose w_T is X-inert in $t_1, \dots t_d$, supported on $t_1 \asymp Z$ and $t_i \asymp X_i$ for $i = 2, \dots, d$. Suppose that on the support of w_T, $\phi = \phi_T$ satisfies*

$$\frac{\partial^{a_1+a_2+\cdots+a_d}}{\partial t_1^{a_1} \dots \partial t_d^{a_d}} \phi(t_1, t_2, \dots, t_d) \ll_{C_{\mathcal{F}}} \frac{Y}{Z^{a_1}} \frac{1}{X_2^{a_2} \dots X_d^{a_d}}, \tag{132}$$

for all $a_1, \dots, a_d \in \mathbb{Z}_{\geq 0}$. Suppose $\phi''(t_1, t_2, \dots, t_d) \gg \frac{Y}{Z^2}$, (here and below, ϕ' and ϕ'' denote the derivative with respect to t_1) for all $t_1, t_2, \dots, t_d$ in the support of w_T, and there exists $t_0 \in \mathbb{R}$ such that $\phi'(t_0, t_2, \dots, t_d) = 0$ with some $t_0 \in \mathbb{R}$ depending on $t_2, \dots, t_d$ (note t_0 is necessarily unique). Suppose that $Y/X^2 \geq R \geq 1$. Then

$$I = \int_{\mathbb{R}} e^{i\phi(t_1,\dots,t_d)} w_T(t_1, \dots, t_d) dt_1 = \frac{Z}{\sqrt{Y}} e^{i\phi(t_0,t_2,\dots,t_d)} W_T(t_2, \dots, t_d) + O_A(ZR^{-A}), \tag{133}$$

for some X-inert family of functions W_T, and where $A > 0$ is arbitrarily large. The implied constant in (133) depends only on A and $C_{\mathcal{F}}$.

10.4 The integral transform

Here we obtain useful expressions for K, which was defined in (92). The key is not an exact formula for K, but rather a Mellin formula with the variables separated. Throughout the remainder of this section, w_T will denote a member of a 1-inert family of functions, which may change from line-to-line without explicit mention. We also recall that $[c, q] = c\frac{q_e}{(c,q_e)}$, where $\frac{q_e}{(c,q_e)}$ takes the possible values 1, 2, 4.

Lemma 11 (Oscillatory case) *Suppose that* $|m_i| \asymp M_i$ *for* $i = 1, 2, 3$, *and* $c \asymp C$, *with* $\tilde{q}|c$. *Suppose that*

$$\frac{\sqrt{ABN_1N_2N_3}}{C} \gg (qr)^{\varepsilon}. \tag{134}$$

Then

$$K(m_1, m_2, m_3, c) = \frac{C^{3/2}(N_1N_2N_3)^{1/2} e\left(\frac{-m_1m_2m_3c^2}{AB[c,q]^3}\right)}{(M_1M_2M_3)^{1/2}} L(m_1, m_2, m_3, c) + O((qr)^{-1/\varepsilon} \prod_{i=1}^{3} (1 + |m_i|)^{-2}), \tag{135}$$

where L *has the following properties. Firstly,* L *vanishes (meaning* K *is very small) unless*

$$M_i \asymp \frac{(ABN_1N_2N_3)^{1/2}}{N_i}, \quad i = 1, 2, 3, \tag{136}$$

and all the m_i *have the same sign. Moreover, we have with*

$$P = \frac{M_1M_2M_3}{ABC}. \tag{137}$$

that

$$L(m_1, m_2, m_3, c) = \frac{1}{P^{1/2}} \int_{|\mathbf{u}| \ll (qr)^{\varepsilon}} \int_{|y| \ll (qr)^{\varepsilon}} F(\mathbf{u}; y) \left(\frac{|m_1m_2m_3|c^2}{[c,q]^3}\right)^{iy} \times \left(\frac{M_1}{|m_1|}\right)^{u_1} \left(\frac{M_2}{|m_2|}\right)^{u_2} \left(\frac{M_3}{|m_3|}\right)^{u_3} \left(\frac{C}{c}\right)^{u_4} d\mathbf{u}dy, \tag{138}$$

where $F = F_{A,B,C,N_1,N_2,N_3,M_1,M_2,M_3}$ *is entire in terms of* $\mathbf{u}$, *and satisfies* $F(\mathbf{u}; y) \ll_{Re(\mathbf{u})} (1 + |\mathbf{u}|)^{-J}(1 + |y|)^{-J}$, *for* J *arbitrarily large. Here* F *additionally depends on the choice of signs of the* m_i, *and on the values* q_e, (c, q_e).

Lemma 12 (Non-oscillatory case) *Suppose that* $|m_i| \asymp M_i$ *for* $i = 1, 2, 3$, $c \asymp C$, *and*

$$\frac{\sqrt{ABN_1N_2N_3}}{C} \ll (qr)^{\varepsilon}. \tag{139}$$

Then

$$K(m_1, m_2, m_3, c) = N_1N_2N_3 \left(\frac{\sqrt{ABN_1N_2N_3}}{C}\right)^{\kappa-1} e\left(\frac{-m_1m_2m_3c^2}{AB[c,q]^3}\right) \int_{|\mathbf{u}| \ll (qr)^{\varepsilon}} F(\mathbf{u}) \int_{|t| \ll (qr)^{\varepsilon}+P}$$

$$\times f(t)\left(\frac{|m_1m_2m_3|c^2}{[c,q]^3}\right)^{it}\left(\frac{M_1}{|m_1|}\right)^{u_1}\left(\frac{M_2}{|m_2|}\right)^{u_2}\left(\frac{M_3}{|m_3|}\right)^{u_3}\left(\frac{C}{c}\right)^{u_4}dtd\mathbf{u}$$
$$+O((qr)^{-1/\varepsilon}\prod_{i=1}^{3}(1+|m_i|)^{-2}), \tag{140}$$

where P is given by (137), $f(t) \ll (1+|t|)^{-1/2}$, and $F(\mathbf{u}) \ll_{J,Re(\mathbf{u})} \prod_{\ell=1}^{4}(1+\frac{(qr)^{\varepsilon}}{|u_\ell|})^{-J}$. Moreover, f vanishes (meaning, K is small) unless

$$\frac{M_1N_1}{C} \ll (qr)^{\varepsilon}, \quad \frac{M_2N_2}{C} \ll (qr)^{\varepsilon}, \quad \frac{M_3N_3}{C} \ll (qr)^{\varepsilon}. \tag{141}$$

If $P \gg (qr)^{\varepsilon}$, the function f may be chosen to have support on $|t| \asymp P$.

Lemma 13 (Other cases) *Suppose some $m_i = 0$. If (134) holds, then K is small. If (139) holds, then K is small unless $|m_j| \ll \frac{C}{N_j}(qr)^{\varepsilon}$ for $j = 1, 2, 3$, in which case*

$$K(m_1,m_2,m_3;c) \ll \left(\frac{\sqrt{ABN_1N_2N_3}}{C}\right)^{\kappa-1}N_1N_2N_3. \tag{142}$$

If say $m_3 = 0$ but $m_1m_2 \neq 0$, then with $N = N_1N_2N_3$, we have a Mellin formula

$$K(m_1,m_2,0;c) = \left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1}N\int_{|v_1|\ll(qr)^{\varepsilon}}\int_{|v_2|\ll(qr)^{\varepsilon}}\left(\frac{C}{N_1|m_1|}\right)^{v_1}$$
$$\times\left(\frac{C}{N_2|m_2|}\right)^{v_2}R(v_1,v_2,c)dv_1dv_2 + O((qr)^{-1/\varepsilon}\prod_{i=1}^{3}(1+|m_i|)^{-2}), \tag{143}$$

where $R(v_1,v_2,c)$ is analytic in $Re(v_i) > 0$ for $i = 1, 2$, and satisfies the bound

$$R(v_1,v_2,c) \ll_{J,Re(v_1),Re(v_2)} \prod_{j=1}^{2}\left(1+\frac{(qr)^{\varepsilon}}{|v_j|}\right)^{-J}.$$

Here R depends on the choices of sign of the m_i, but we suppress it from the notation. Similar formulas hold when $m_1 = 0$ or $m_2 = 0$.

If two $m_i = 0$ but the other m_i is nonzero, then a formula similar to (143) holds, but with one of the integrals omitted.

Proof We prove all three lemmas.

Truncations As our first step, we integrate by parts three times in each of the t_i for which $m_i \neq 0$, allowing us to obtain a crude bound of the form

$$K(m_1,m_2,m_3,c) \ll P(q,r,N_1,N_2,N_3,c)\prod_{i=1}^{3}(1+|m_i|)^{-3},$$

where P is some fixed polyomial (which may as well be taken to be some polynomial in qr by (88)). This bound is sufficient for the lemmas when some $|m_i|$ is $\gg (qr)^{A'}$ for some large A' depending polynomially on $1/\varepsilon$. For the rest of the proof, suppose that $|m_i| \ll (qr)^{A'}$ for some A', and each i.

In the oscillatory case (i.e., if (134) holds), then using the fact that $J_{\kappa-1}(x) = e^{ix}W_+(x) + e^{-ix}W_-(x)$ where $W_\pm(x)$ have controlled derivatives (cf. Watson [44, Page 205]), we see that repeated integration by parts (Lemma 10) shows that K is very small unless (136) holds, and moreover, all the m_i must have the same sign. Similarly, in the non-oscillatory case where (139) holds, then repeated integration by parts shows that K is small unless (141) holds. □

Proof of Lemma 11 Now we show the expression (135), with L given by (138). Using the Fourier integral (valid for n an odd integer)

$$J_n(x) = \sum_{\pm} \frac{\pm 1}{\pi i} \int_0^{\pi/2} \sin(nv) e^{\pm ix \sin v} dv,$$

we have

$$\begin{aligned} K &= K(m_1, m_2, m_3; c) \\ &= \sum_{\pm} \frac{\pm 1}{\pi i} \int_0^{\pi/2} \sin((\kappa-1)v) \int_{\mathbb{R}^3} e\left(\frac{\pm 2\sqrt{ABt_1t_2t_3}}{c} \sin(v)\right) \\ &\quad \times e\left(\frac{-m_1t_1 - m_2t_2 - m_3t_3}{[c,q]}\right) w_T(t_1, t_2, t_3) dt_1 dt_2 dt_3 dv. \end{aligned}$$

Next we change variables $t_3 = u\frac{N_1N_2}{t_1t_2}$, giving

$$K = \sum_{\pm} \frac{\pm 1}{\pi i} \int_0^{\pi/2} \sin((\kappa-1)v) \int_0^{\infty} e\left(\frac{\pm 2\sqrt{ABuN_1N_2}}{c} \sin(v)\right) I(u) du dv,$$

where

$$I(u) = \int_{\mathbb{R}^2} e\left(\frac{-m_1t_1 - m_2t_2 - m_3\frac{uN_1N_2}{t_1t_2}}{[c,q]}\right) w_T(t_1, t_2, u) dt_1 dt_2.$$

The conditions (134) and (136) imply that

$$\frac{M_iN_i}{C} \asymp \frac{\sqrt{ABN_1N_2N_3}}{C} \gg (qr)^{\varepsilon}.$$

Recall that we already showed, in the paragraphs following the statement of Lemma 13, that under the assumption of (134), K is small unless (136) holds. The conditions are now in place to analyze the inner t_1, t_2 integrals using stationary phase. Evaluating the t_1-integral first, we find a stationary point at $t_1^0 = (\frac{m_3uN_1N_2}{m_1t_2})^{1/2}$. Following this, the

stationary point in terms of t_2 occurs at $t_2^0 = m_2^{-2/3}(m_1m_3uN_1N_2)^{1/3}$. We therefore deduce

$$I(u) = \frac{C}{(N_1N_2M_1M_2)^{1/2}} e\left(\frac{-3(m_1m_2m_3uN_1N_2)^{1/3}}{[c,q]}\right) w_T(u, m_1, m_2, m_3, c) + O((qr)^{-1/\varepsilon}\prod_{i=1}^{3}(1+|m_i|)^{-2}), \tag{144}$$

where w_T is 1-inert in all variables and T stands for the tuple

$$(M_1, M_2, M_3, N_1, N_2, N_3, C, q_e, (c, q_e))$$

and supported on $u \asymp N_3$. The condition that m_1, m_2, m_3 must all have the same sign may be encoded in the support of the inert function. If all three terms are negative, we naturally interpret the expression $(m_1m_2m_3)^{1/3}$ as $-(|m_1m_2m_3|)^{1/3}$. Therefore

$$K = \left[\sum_{\pm}\frac{\pm 1}{\pi i}\frac{C}{(N_1N_2M_1M_2)^{1/2}}\int_0^{\pi/2}\sin((\kappa-1)v) \times \int_0^{\infty} e\left(\frac{\pm 2\sqrt{ABuN_1N_2}}{c}\sin(v)\right) e\left(\frac{-3(um_1m_2m_3N_1N_2)^{1/3}}{[c,q]}\right) w_T(u,\cdot)dudv\right] + O\left((qr)^{-1/\varepsilon}\prod_{i=1}^{3}(1+|m_i|)^{-2}\right). \tag{145}$$

Here we use the notation $w_T(u, \cdot)$ to denote a function where we are currently focusing on the variable u only, and so do not display the other variables.

Finally, we study

$$K_0(v) := \frac{C}{(N_1N_2M_1M_2)^{1/2}}\int_0^{\infty} e\left(\frac{\pm 2\sqrt{ABuN_1N_2}}{c}\sin(v)\right) \times e\left(\frac{-3(um_1m_2m_3N_1N_2)^{1/3}}{[c,q]}\right) w_T(u,\cdot)du.$$

We will presently show that

$$K_0(v) = \frac{C^{3/2}(N_1N_2N_3)^{1/2}}{(M_1M_2M_3)^{1/2}} e\left(\frac{-m_1m_2m_3c^2}{AB[c,q]^3\sin^2 v}\right) w_T(\sin v, \cdot) + O((qr)^{-1/\varepsilon}\prod_{i=1}^{3}(1+|m_i|)^{-2}), \tag{146}$$

where $w_T(\sin v, \cdot)$ is part of a 1-inert family of functions of $\sin v, m_1, m_2, m_3, c$, with T as before but in addition depending on the choice of $\pm$ sign.

This integral defining $K_0(v)$ is small unless there is a stationary point (by Lemma 10 again), which occurs at

$$u = u_0 = \frac{(m_1 m_2 m_3)^2 c^6}{N_1 N_2 (AB)^3 \sin^6 v [c,q]^6},$$

under the additional assumption that the choice of $\pm$ sign matches the sign of $m_1 m_2 m_3$ (which in turn has the same sign as each individual m_i). Note that for this stationary point to lie inside the support of the inert function, we need

$$\sin v \asymp \frac{(M_1 M_2 M_3)^{1/3}}{(AB)^{1/2}(N_1 N_2 N_3)^{1/6}} =: V. \tag{147}$$

Thus we obtain, in both cases of $\pm$ sign, that

$$K_0(v) = (\text{scaling factor}) e\left(\frac{-m_1 m_2 m_3 c^2}{AB \sin^2 v [c,q]^3}\right) w_T(\sin v, \cdot) + O((qr)^{-1/\varepsilon} \prod_{i=1}^{3} (1+|m_i|)^{-2}).$$

A short calculation shows the scaling factor is $\asymp \frac{C^{3/2}(N_1 N_2 N_3)^{1/2}}{(M_1 M_2 M_3)^{1/2}}$, proving the claimed expression for $K_0(v)$.

Thus we obtain

$$K = \frac{C^{3/2}(N_1 N_2 N_3)^{1/2}}{(M_1 M_2 M_3)^{1/2}} \int_0^{\pi/2} \sin((\kappa-1)v) e\left(\frac{-m_1 m_2 m_3 c^2}{AB \sin^2 v [c,q]^3}\right) w_T(\sin v, \cdot) dv,$$

plus an error of size $O((qr)^{-1/\varepsilon} \prod_i (1+|m_i|)^{-2})$, where the support of the inert function is given by (147).

Next we factor out the desired exponential, giving now

$$K = e\left(\frac{-m_1 m_2 m_3 c^2}{AB[c,q]^3}\right) \frac{C^{3/2}(N_1 N_2 N_3)^{1/2}}{(M_1 M_2 M_3)^{1/2}} \times \int_0^{\pi/2} \sin((\kappa-1)v) e\left(\frac{m_1 m_2 m_3 c^2}{AB[c,q]^3}\left(1 - \frac{1}{\sin^2 v}\right)\right) w_T(\sin v, \cdot) dv,$$

plus a small error term. Define

$$K_{00}(x) := \int_0^{\pi/2} \sin((\kappa-1)v) e\left(x\left(-1 + \frac{1}{\sin^2 v}\right)\right) w_T(\sin v, \cdot) dv. \tag{148}$$

Note that for our particular values of the parameters, we have

$$\frac{x}{V^2} \gg (qr)^{\varepsilon}.$$

Now we asymptotically evaluate $K_{00}(x)$. First we dispense with the case where the support of the inert function is so that the integrand vanishes unless $v \leq \frac{\pi}{2} - \frac{\pi}{100}$, say. Thus $\cos(v) \asymp 1$ and $\sin(v) \asymp v \asymp V$ in the support of the integrand. We claim that $K_{00}(x) \ll (qr)^{-A}$ in this case. To see this, we first note that it suffices to bound

$$\int_{-\infty}^{\infty} w_T(\sin v, \cdot) e(\phi(v)) dv, \quad \text{where} \quad \phi(v) = \pm \frac{\kappa - 1}{2\pi} v + x \frac{\cos^2 v}{\sin^2 v}.$$

We have

$$\phi(v) = \pm \frac{\kappa - 1}{2\pi} v + \frac{x}{v^2} \left(1 + c_2 v^2 + c_4 v^4 + \cdots\right),$$

for certain constants c_i. Then we have

$$\phi'(v) \asymp \frac{x/v^2}{v}, \qquad \phi^{(j)}(v) \ll \frac{x/v^2}{v^j}, \quad j = 2, 3, \ldots,$$

using $\frac{x}{v^3} \gg \frac{x}{v^2} \gg (qr)^{\varepsilon} \gg \kappa - 1$, since κ is fixed. By Lemma 10 yet again, the integral is very small. If the inert function has support on an interval containing $\pi/2$, then the above argument breaks down. So now suppose that the inert function has support on $v \geq \pi/4$, so in particular $V \asymp 1$ and $x \gg (qr)^{\varepsilon}$. Change variables $v = \pi/2 - u$, giving

$$K_{00}(x) = (-1)^{\frac{\kappa-2}{2}} \int_0^{\pi/4} \cos((\kappa - 1)u) e\left(x \frac{\sin^2 u}{\cos^2 u}\right) w_T(\cos u, \cdot) du. \tag{149}$$

Next we argue that the main part of this integral comes from $u \ll x^{-1/2}(qr)^{\varepsilon}$, provided we use a smooth truncation. Let us apply a partition of unity, and consider the portion of the integral with $u \asymp U$ (with $U \ll 1$), which we denote $K_{00}^U(x)$. By Lemma 10 yet again, if $xU^2 \gg (qr)^{\varepsilon}$, then $K_{00}^U(x)$ is small. We may also use that the integrand is even to extend to $-\pi/4$ to $\pi/4$, giving

$$K_{00}(x) = \int_{-\pi/4}^{\pi/4} W(u) e(x \tan^2 u) du,$$

plus a small error term, where W has support on $|u| \ll U$ with $U = x^{-1/2+\varepsilon}$.

We may now derive an asymptotic expansion of K_{00}, with leading term $c_0 W(0) x^{-1/2}$, where c_0 is some absolute constant. By developing this expansion carefully, we have that for $x \gg (qr)^{\varepsilon}$, $K_{00}^{(j)}(x) \ll x^{-1/2-j}$, and so by Mellin inversion, we have that

$$K_{00}(x) = x^{-1/2} \int_{-\infty}^{\infty} f(t) x^{it} dt,$$

plus a small error term, where $|f(t)| \ll_A (1 + |t|)^{-A}$, with $A > 0$ arbitrarily large. In our application, we may thus truncate at $|t| \ll (qr)^{\varepsilon}$.

The previous discussion gives a formula for K of the form (135), where L takes the form

$$L(m_1, m_2, m_3, c) = \frac{1}{P^{1/2}} \int_{|t| \ll (qr)^\varepsilon} w_T(t, m_1, m_2, m_3, c) \left(\frac{|m_1 m_2 m_3| c^2}{[c, q]^3} \right)^{it} dt,$$

where w_T is 1-inert in the variables m_1, m_2, m_3, c and has rapid decay in t, uniformly in all other parameters. We may then write

$$w_T(t, m_1, m_2, m_3, c) = \int F(\mathbf{u}; t) \left(\frac{M_1}{|m_1|} \right)^{u_1} \left(\frac{M_2}{|m_2|} \right)^{u_2} \left(\frac{M_3}{|m_3|} \right)^{u_3} \left(\frac{C}{c} \right)^{u_4} d\mathbf{u},$$

where the integral is over four vertical lines in the complex plane, one for each of the u_i, $i = 1, 2, 3, 4$. By the rapid decay of F beyond $(qr)^\varepsilon$, due to the fact that w_T is inert, we may truncate the integrals at $|\mathbf{u}| \ll (qr)^\varepsilon$. This expression gives (138), and so completes the proof of Lemma 11. □

Proof of Lemma 12 Now suppose (139) holds. As previously mentioned in the paragraphs following Lemma 13, K is small unless (141) holds, a condition that we assume henceforth. Assuming $x \ll X = 1 + \frac{\sqrt{ABN_1N_2N_3}}{C}$, we have that $J_{\kappa-1}(x) = x^{\kappa-1} W(x)$ where W is a smooth function satisfying $x^j W^{(j)}(x) \ll X^j$. That is, W satisfies the same derivative bounds as an X-inert function, and so it may be absorbed into the definition of the inert function. Therefore, by the separation of variables discussion from Sect. 10.1, we have

$$\begin{aligned} K = N_1 N_2 N_3 \left(\frac{\sqrt{ABN_1N_2N_3}}{C} \right)^{\kappa-1} \int_{|\mathbf{u}| \ll (qr)^\varepsilon} F(\mathbf{u}) \left(\frac{M_1}{|m_1|} \right)^{u_1} \left(\frac{M_2}{|m_2|} \right)^{u_2} \\ \times \left(\frac{M_3}{|m_3|} \right)^{u_3} \left(\frac{C}{c} \right)^{u_4} d\mathbf{u} + O((qr)^{-A} \prod_{i=1}^{3} (1 + |m_i|)^{-2}), \end{aligned} \tag{150}$$

where $F(\mathbf{u}) \ll_{J, \mathrm{Re}(\mathbf{u})} \prod_{\ell=1}^{4} (1 + \frac{X}{|u_\ell|})^{-J}$.

We also want to factor out the exponential term $e_{AB[c,q]^3/c^2}(-m_1 m_2 m_3)$. It is not clear whether

$$P := \frac{M_1 M_2 M_3}{ABC}$$

is $\gg 1$ or $\ll 1$, so we treat both cases separately.

If $P \ll (qr)^\varepsilon$, then essentially the exponential term $e_{AB[c,q]^3/c^2}(m_1 m_2 m_3)$ is not oscillatory, so by Mellin inversion there exists a simple formula of the form

$$\begin{aligned} e\left(\frac{m_1 m_2 m_3 c^2}{AB[c, q]^3} \right) w\left(\pm \frac{m_1 m_2 m_3 c^2 / (AB[c, q]^3)}{P} \right) \\ = \int_{|t| \ll (qr)^\varepsilon} \left(\frac{|m_1 m_2 m_3| c^2}{[c, q]^3} \right)^{it} f(t) dt + O((qr)^{-1/\varepsilon}), \end{aligned} \tag{151}$$

where $f_{A,B,C,M_1,M_2,M_3,\pm}(t) = f(t) \ll 1$ and $w(t)$ is a smooth function of compact support on $(0,\infty)$, that is identically 1 on the support of the inert function w_T in the definition of K.

If $P \gg (qr)^\varepsilon$, then we claim that a formula like (151) holds, but with $|t| \asymp P$ and $f(t) \ll |t|^{-1/2}$. For this, we argue as follows. By Mellin inversion, we have $e^{ix}w(x/P) = \int_{-\infty}^{\infty} f(t)x^{it}dt$, where $f(t) = \frac{1}{2\pi}\int_0^\infty e^{ix}w(x/P)x^{it-1}dx$. Since the support on w causes $x \asymp P$, repeated integration by parts shows $f(t)$ is very small except when $|t| \asymp P$. A standard stationary phase bound gives $f(t) \ll |t|^{-1/2}$.

In either case, we obtain (140).

Proof of Lemma 13 The claims that if (134) holds, then K is small, and that if (139) holds, K is small unless $|m_i| \ll \frac{C}{N_i}(qr)^\varepsilon$ for $i = 1, 2, 3$ have already been shown in the previous analysis. It remains to show the integral formula.

Suppose that $m_3 = 0, m_1, m_2 \neq 0$, and (139) holds. The idea is to apply an analog of (150), but only in the m_1, m_2 variables. This case is easier because the exponential factor simplifies as 1, so there is no need to separate out $e_{AB[c,q]^3}(m_1m_2m_3c^2)$. By taking a Mellin transform in the m_1, m_2 variables, we get

$$\begin{aligned} &K(m_1, m_2, 0; c) \\ &= \left(\frac{\sqrt{ABN_1N_2N_3}}{C}\right)^{\kappa-1} N_1N_2N_3 \int_{|v_1|\ll(qr)^\varepsilon}\int_{|v_2|\ll(qr)^\varepsilon} \left(\frac{C}{N_1|m_1|}\right)^{v_1}\left(\frac{C}{N_2|m_2|}\right)^{v_2} \\ &\quad \times R(v_1, v_2, c)dv_1dv_2 + O\left(q^{-A}\prod_{i=1}^{3}(1+|m_i|)^{-2}\right), \end{aligned}$$

where $R(v_1, v_2, c)$ is analytic in $\mathrm{Re}(v_i) > 0$ for $i = 1, 2$, and satisfies the bound

$$R(v_1, v_2, c) \ll_{J,\varepsilon,\mathrm{Re}(v_1),\mathrm{Re}(v_2)} \prod_{j=1}^{2}\left(1 + \frac{(qr)^\varepsilon}{|v_j|}\right)^{-J}.$$

This bound on R is precisely analogous to the bound on $F(\mathbf{u})$ in the proof of Lemma 12. This completes the proof. □

11 Recombination

Now we prove Proposition 2. Recall formulas (87), (89), and (90). Let us write $\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_1$ where $\mathcal{S}_0$ corresponds to the terms with some $m_i = 0$, while $\mathcal{S}_1$ corresponds to the terms with all $m_i \neq 0$.

11.1 Bounding $\mathcal{S}_0$

Lemma 14 *We have* $\mathcal{S}_0 \ll R^{-1}(qr)^\varepsilon$.

Proof From Lemma 13, we see that K, and hence $\mathcal{S}_0$, is small unless (139) holds. By (139) and (88) we have that A, B, and C are each bounded by $(qr)^{10}$, say. This allows us to replace several factors of $(AB)^{\varepsilon}$ and C^{ε} by $(qr)^{\varepsilon}$ in the following.

Let us further decompose $\mathcal{S}_0 = \mathcal{S}_{00} + \mathcal{S}_{01} + \mathcal{S}_{02}$ where $\mathcal{S}_{02}$ corresponds to the terms with exactly two of the $m_i \neq 0$, $\mathcal{S}_{01}$ corresponds to the terms with precisely one of the $m_i \neq 0$, and $\mathcal{S}_{00}$ corresponds to the terms with $m_1 = m_2 = m_3 = 0$.

We first bound $\mathcal{S}_{02}$. Suppose for the sake of argument that $m_3 = 0$, and $m_1, m_2 \geq 1$, the other cases being similar, and let $\mathcal{S}_{02}^{+}$ denote these terms. For $i = 1, 2$, let $m_i = m_i' m_i''$ where $(m_i', 2AB) = 1$ and $m_i'' \mid (2AB)^{\infty}$. Then by Lemmas 9 and 13 we have

$$|\mathcal{S}_{02}^{+}| \ll \sum_{N_1, N_2, N_3, C} \frac{\sqrt{N}}{C} \left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1} \times \sum_{\substack{c \equiv 0 \pmod{\tilde{q}R} \\ c \asymp C}} \frac{1}{qc} \left| \sum_{m_1, m_2 \geq 1} g(m_1'', m_2'', 0, A, B, c, q) R_{q_o}(m_1') R_{q_o}(m_2') \right. \times \left. \int_{|v_1| \ll (qr)^{\varepsilon}} \int_{|v_2| \ll (qr)^{\varepsilon}} \left(\frac{C}{N_1 m_1}\right)^{v_1} \left(\frac{C}{N_2 m_2}\right)^{v_2} R(v_1, v_2, c)\, dv_1\, dv_2 \right|, \tag{152}$$

plus a small error term.

The goal is to form Dirichlet series over m_1', m_2'. Since q_o is square-free and coprime to $2AB$, we have that

$$Q_{q_o}(s) := \sum_{(m, 2AB) = 1} \frac{R_{q_o}(m)}{m^s} = \mu(q_o) \zeta^{(2AB)}(s) \prod_{p \mid q_o} (1 - p^{1-s}).$$

Then factoring $m_i = m_i'' m_i'$, and using the bound on g from Lemma 9 we have

$$|\mathcal{S}_{02}^{+}| \ll \sum_{N_1, N_2, N_3, C} \frac{\sqrt{N}}{C} \left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1} \sum_{\substack{c \equiv 0 \pmod{\tilde{q}R} \\ c \asymp C}} \frac{(A, c)(B, c)}{qc} \sum_{m_1'', m_2'' \mid (2AB)^{\infty}} \times \left| \iint_{\substack{|v_1| \ll (qr)^{\varepsilon} \\ |v_2| \ll (qr)^{\varepsilon}}} \left(\frac{C}{N_1 m_1''}\right)^{v_1} \left(\frac{C}{N_2 m_2''}\right)^{v_2} Q_{q_o}(v_1) Q_{q_o}(v_2) R(v_1, v_2, c)\, dv_1\, dv_2 \right|, \tag{153}$$

plus a small error term.

We may assume without loss of generality that $q > 8$, so that $q_o > 1$. Note that under the assumption that $q_o > 1$, the function $Q_{q_o}(s)$ is holomorphic in $\mathrm{Re}(s) > 0$. On the $\mathrm{Re}(s) = \varepsilon$ line, it satisfies $Q_{q_o}(s) \ll (1 + |s|)^{1/2} \phi(q_o) (qr)^{\varepsilon}$. So we can move the vertical contour integrals to the lines $\mathrm{Re}(v_1) = \mathrm{Re}(v_2) = \varepsilon$. The short horizontal segments created in doing so are extremely small by the bounds on $R(v_1, v_2, c)$ from

Lemma 13. From the vertical lines of integration we obtain $S_{02}(N_1, N_2, N_3, c) \ll (qr)^{\varepsilon}\phi(q)^2(A, c)(B, c)$. Therefore we have

$$|\mathcal{S}_{02}^{+}| \ll (qr)^{\varepsilon} \sum_{N_1,N_2,N_3,C} \frac{\sqrt{N}}{C}\left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1} \sum_{\substack{c\equiv 0 \pmod{\tilde{q}R} \\ c\asymp C}} \frac{\phi(q)^2}{qc}(A, c)(B, c), \tag{154}$$

plus a small error term.

Using $(qR, AB) = 1$, Cauchy's inequality, and

$$\sum_{n\le x}(d, n)^2 \ll x\tau(d)d, \tag{155}$$

we derive

$$\sum_{\substack{c\equiv 0 \pmod{\tilde{q}R} \\ c\asymp C}} (A, c)(B, c) \ll \frac{C}{qR}(AB)^{1/2}(qr)^{\varepsilon}.$$

Therefore by (88) and (139)

$$|\mathcal{S}_{02}^{+}| \ll \sum_{N_1,N_2,N_3,C} (qr)^{\varepsilon}\frac{NAB}{C^2R} \ll \frac{(qr)^{\varepsilon}}{R}, \tag{156}$$

which is sufficient for the bound in the statement of the lemma. By a symmetry argument, this shows the desired bound on $\mathcal{S}_{02}$.

Similarly to the method used to bound $\mathcal{S}_{02}$, we have in analogy with (153), the bound

$$|\mathcal{S}_{01}^{+}| \ll \sum_{N_1,N_2,N_3,C} \frac{(qr)^{\varepsilon}}{\sqrt{N}C}\left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1} N \sum_{\substack{c\equiv 0 \pmod{\tilde{q}R} \\ c\asymp C}} \frac{(A, c)(B, c)\phi(q)}{cq}$$
$$\times \sum_{m_1''|(2AB)^{\infty}} \left|\int_{|v|\ll(qr)^{\varepsilon}} \left(\frac{C}{N_1m_1''}\right)^{v} Q_{q_o}(v)R(v, c)dv\right|,$$

where $Q_{q_o}(v)R(v, c)$ is analytic in $\mathrm{Re}(v) > 0$, and satisfies the bound

$$Q_{q_o}(v)R(v, c) \ll_{\mathrm{Re}(v)} (1 + |v|)^{1/2}(qr)^{\varepsilon}\phi(q_o).$$

We move the contour to the line $\mathrm{Re}(v) = \varepsilon$. The short horizontal segment created in doing so is extremely small by the bounds on $R(v, c)$ from Lemma 13. From the vertical segment we get

$$|\mathcal{S}_{01}^{+}| \ll \sum_{N_1,N_2,N_3,C} \frac{(qr)^{\varepsilon}}{\sqrt{N}C}\left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1} N \sum_{\substack{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)\\ c\asymp C}} \frac{(A,c)(B,c)}{cq}\phi(q)^2 \tag{157}$$

plus a small error term. This is precisely the same bound as (154), and so our final bound on $\mathcal{S}_{01}$ is identical to the bound on $\mathcal{S}_{02}$.

Finally, we bound $\mathcal{S}_{00}$, which is the easiest case. Using only the upper bound (142) and the upper bound from Lemma 9, we obtain a bound on $\mathcal{S}_{00}$ of the exact same shape as (154), so the proof is complete. □

11.2 Bounding $\mathcal{S}_1$

Here we show the desired bound

$$\mathcal{S}_1 \ll (qr)^{\varepsilon}\left(\frac{(AB)^{1/2}}{R^{1/2}} + \frac{r^{3/4}}{q^{1/2}R}\right). \tag{158}$$

Let us write $\mathcal{S}_1 = \sum_{\epsilon_1,\epsilon_2,\epsilon_3\in\{\pm\}} \mathcal{S}_1^{\epsilon_1,\epsilon_2,\epsilon_3}$ where this sum is restricted to $\epsilon_i m_i > 0$ for $i = 1, 2, 3$. The same method will apply to each of these terms, so for simplicity we estimate $\mathcal{S}_1^{+,+,+}$, which we denote with shorthand by $\mathcal{S}_1^{+}$.

We have

$$\mathcal{S}_1^{+} = \sum_{\substack{M_1,M_2,M_3\\ N_1,N_2,N_3,C}} \frac{1}{N^{1/2}C} \sum_{m_i\asymp M_i} \sum_{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)} w_C(c) \left(e\left(\frac{-m_1m_2m_3}{AB[c,q]^3/c^2}\right) G_{A,B}(m_1,m_2,m_3;c)\right)\left(e\left(\frac{m_1m_2m_3}{AB[c,q]^3/c^2}\right) K(m_1,m_2,m_3;c)\right).$$

There are two main cases to consider, depending on if (134) holds, or if (139) holds, and we correspondingly write $\mathcal{S}_1^{+} = \mathcal{T} + \mathcal{U}$.

Case $\mathcal{T}$. By Lemma 11, we have (with shorthand $M = M_1M_2M_3$, and M_i satisfying (136))

$$\mathcal{T} \ll \sum_{\substack{N_1,N_2,N_3,C\\ M_1,M_2,M_3}} \Bigg| \frac{C^{3/2}N^{1/2}}{C^2q\sqrt{MN}} \sum_{\substack{c\equiv 0 (\mathrm{mod}\ \tilde{q}R)\\ c\asymp C}} \times \sum_{m_1,m_2,m_3} \left(e\left(\frac{-m_1m_2m_3}{AB[c,q]^3/c^2}\right) cqG_{A,B}(m_1,m_2,m_3;c)\right) \frac{1}{P^{1/2}} \int_{|\mathbf{u}|\ll(qr)^{\varepsilon}} \int_{|y|\ll(qr)^{\varepsilon}} \times F(\mathbf{u};y)\left(\frac{m_1m_2m_3c^2}{[c,q]^3}\right)^{iy}\left(\frac{M_1}{m_1}\right)^{u_1}\left(\frac{M_2}{m_2}\right)^{u_2}\left(\frac{M_3}{m_3}\right)^{u_3}\left(\frac{C}{c}\right)^{u_4} d\mathbf{u}dy \Bigg|,$$

plus a small error term. Assume that $\mathrm{Re}(u_i) = 1 + \varepsilon$ for all $i = 1, 2, 3, 4$. Then by (93), we have

$$\mathcal{T} \ll \sum_{N_1,N_2,N_3,C} \Bigg| \frac{C^{-1/2}}{q(PM)^{1/2}} \int_{|\mathbf{u}|\ll (qr)^\varepsilon} \int_{|y|\ll (qr)^\varepsilon} F(\mathbf{u}; y) \sum_{\substack{(\epsilon_1,\epsilon_2,\epsilon_3)\in\{\pm 1\}^3 \\ \delta\in\{1,2,4,8\}}} \left(\frac{\delta^3}{\tilde{q} R q_e^3}\right)^{iy} \times Z_{\delta,R,q}^{(\epsilon_1,\epsilon_2,\epsilon_3)}(u_1 - iy, u_2 - iy, u_3 - iy, u_4 + iy) M_1^{u_1} M_2^{u_2} M_3^{u_3} \left(\frac{C}{\tilde{q}R}\right)^{u_4} d\mathbf{u} dy \Bigg|,$$

plus a small error term. Now we decompose further by $Z = Z_0 + Z'$, as in Proposition 3, and write $\mathcal{T} = \mathcal{T}_0 + \mathcal{T}'$. For $\mathcal{T}_0$ we have

$$\mathcal{T}_0 \ll \frac{C^{1/2}}{(PM)^{1/2}} \frac{M}{q^2 RAB} (qr)^\varepsilon.$$

Recall $P = \frac{M}{ABC}$, $M \asymp (AB)^{3/2} N^{1/2}$, $C \ll \sqrt{NAB}(qr)^\varepsilon$, and $N \ll (q^2 r)^{3/2+\varepsilon}$ giving

$$\mathcal{T}_0 \ll \frac{C}{q^2 R (AB)^{1/2}} (qr)^\varepsilon \ll \frac{N^{1/2}}{q^2 R} (qr)^\varepsilon \ll \frac{1}{\sqrt{q}} \frac{r^{3/4}}{R} (qr)^\varepsilon. \tag{159}$$

For $\mathcal{T}'$, we move the contours to $\mathrm{Re}(u_i) = 1/2 + \varepsilon$ for $i = 1, 2, 3, 4$. The short horizontal segments created in doing so are extremely small by the bounds on $F(\mathbf{u}; y)$ from Lemma 11. Then we obtain

$$\begin{aligned} \mathcal{T}' &\ll (qr)^\varepsilon \sum_{N_1,N_2,N_3,C} \frac{C^{-1/2}}{q(PM)^{1/2}} q^{3/2} \sqrt{AB} M^{1/2} \left(\frac{C}{qR}\right)^{1/2} \\ &\ll (qr)^\varepsilon \sum_{N_1,N_2,N_3,C} \frac{\sqrt{AB}}{P^{1/2} R^{1/2}}. \end{aligned}$$

Since $P \gg 1$, we have

$$\mathcal{T}' \ll (qr)^\varepsilon \frac{(AB)^{1/2}}{R^{1/2}},$$

as desired.

Case $\mathcal{U}$. Here we obtain from Lemma 12 the bound

$$\mathcal{U} \ll \sum_{\substack{N_1,N_2,N_3,C\\ M_1,M_2,M_3}} \Bigg| \frac{N\left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1}}{C^2 q\sqrt{N}} \int_{|\mathbf{u}|\ll(qr)^\varepsilon} \int_{|y|\ll(qr)^\varepsilon+P} F(\mathbf{u}) f_P(y) \times \sum_{\substack{(\epsilon_1,\epsilon_2,\epsilon_3)\in\{\pm1\}^3\\ \delta\in\{1,2,4,8\}}} \left(\frac{\delta^3}{\tilde{q}Rq_e^3}\right)^{iy} Z_{\delta,R,q}^{(\epsilon_1,\epsilon_2,\epsilon_3)}(u_1-iy,u_2-iy,u_3-iy,u_4+iy) M_1^{u_1}M_2^{u_2}M_3^{u_3}\left(\frac{C}{\tilde{q}R}\right)^{u_4} d\mathbf{u}dy \Bigg|, \tag{160}$$

plus a small error term. As in the case of $\mathcal{T}$, write $\mathcal{U} = \mathcal{U}_0 + \mathcal{U}'$. Consider first the case $\mathcal{U}'$. As in case $\mathcal{T}$, we move the contours to $\mathrm{Re}(u_i) = 1/2 + \varepsilon$. The short horizontal segments created in doing so are extremely small by the bounds on $F(\mathbf{u}; y)$ from Lemma 12. Using (96), we get

$$\begin{aligned}\mathcal{U}' &\ll (qr)^\varepsilon \frac{N\sqrt{AB}}{C^3 q}\sqrt{AB}q^{3/2}M^{1/2}\left(\frac{C}{qR}\right)^{1/2}(1+P^{1/2})\\ &\ll (qr)^\varepsilon \frac{NABM^{1/2}}{R^{1/2}C^{5/2}}\left(1+\left(\frac{M}{ABC}\right)^{1/2}\right).\end{aligned}$$

Now $M \ll \frac{C^3}{N}(qr)^\varepsilon$, and (139) holds, so after simplification this leads to

$$\mathcal{U}' \ll (AB)^{1/2}R^{-1/2}(qr)^\varepsilon,$$

as desired.

Finally, we turn to the case of $\mathcal{U}_0$. To start, we suppose that $\mathrm{Re}(u_i) = 1 + \varepsilon$ for $i = 1, 2, 3, 4$.

Consider the case where $P \gg (qr)^\varepsilon$ with sufficiently large implied constant. Then we shift contours to $\mathrm{Re}(u_i) = 1/2 + \varepsilon$, for all i. No poles are encountered in doing so, since $|u_i| < |y|$ throughout the integral (160) if $f_P(y)$ has support $|y| \asymp P$. The short horizontal segments created in preforming this contour shift are extremely small by the bounds on $F(\mathbf{u})$ from Lemma 12. The contribution to $\mathcal{U}_0$ of the integral along the vertical segments is certainly bounded by the same bound we obtained on $\mathcal{U}'$, since the bound on Z_0 appearing in Proposition 3 is much stronger than the bound on Z'. Therefore, $\mathcal{U}_0$ is bounded in a satisfactory way for $P \gg (qr)^\varepsilon$.

Now suppose $P \ll (qr)^\varepsilon$. Then by (94), we have

$$\begin{aligned}\mathcal{U}_0 &\ll (qr)^\varepsilon \sum_{\substack{N_1,N_2,N_3,C\\ M_1,M_2,M_3}} \frac{N\left(\frac{\sqrt{ABN}}{C}\right)^{\kappa-1}}{C^2q\sqrt{N}}\frac{MC}{qRAB}\\ &\ll (qr)^\varepsilon \sum_{\substack{N_1,N_2,N_3,C\\ M_1,M_2,M_3}} N^{1/2}\left(\frac{\sqrt{ABN}}{C}\right)\frac{M}{ABCq^2R}.\end{aligned} \tag{161}$$

Since $P = \frac{M}{ABC} \ll (qr)^{\varepsilon}$ now, and $\frac{\sqrt{ABN}}{C} \ll (qr)^{\varepsilon}$ too, we obtain

$$\mathcal{U}_0 \ll (qr)^{\varepsilon} N^{1/2} q^{-2} R^{-1} \ll (qr)^{\varepsilon} (q^3 r^{3/2})^{1/2} q^{-2} R^{-1} \ll (qr)^{\varepsilon} q^{-1/2} \frac{r^{3/4}}{R}.$$

This is the same bound as (159), which completes the proof of (158).

References

1. Abbes, A., Ullmo, E.: Comparaison des métriques d'Arakelov et de Poincaré sur $X_0(N)$. Duke Math. J. **80**(2), 295–307 (1995)
2. Atkin, A.O.L., Lehner, J.: Hecke operators on $\Gamma_0(m)$. Math. Ann. **185**, 134–160 (1970)
3. Atkin, A.O.L., Li, W.: Twists of newforms and pseudo-eigenvalues of W-operators. Invent. Math. **48**, 221–243 (1978)
4. Barrett, O., Burkhardt, P., DeWitt, J., Dorward, R., Miller, S.J.: One-Level density for holomorphic cusp forms of arbitrary level. Res. Number Theory **3**, 21 (2017) (**Art. 25**)
5. Baruch, E., Mao, Z.: Central value of automorphic L-functions. Geom. Funct. Anal. **17**(2), 333–384 (2007)
6. Blomer, V.: Uniform bounds for Fourier coefficients of theta-series with arithmetic applications. Acta Arith. **114**(1), 1–21 (2004)
7. Blomer, V.: Ternary Quadratic Forms, and Sums of Three Squares with Restricted Variables. Anatomy of Integers, CRM Proc. Lecture Notes, vol. 46. American Mathematical Society, Providence, RI, pp. 1–17 (2008)
8. Blomer, V., Harcos, G.: Hybrid bounds for twisted L-functions. J. Reine Angew. Math. **621**, 53–79 (2008) [**Addendum: Hybrid bounds for twisted L-functions. J. Reine Angew. Math. 694, 241–244 (2014)**]
9. Blomer, V., Harcos, G., Michel, P.: A Burgess-like subconvex bound for twisted L-functions. Appendix 2 by Z. Mao. Forum Math. **19**(1), 61–105 (2007)
10. Blomer, V., Khan, R., Young, M.: Distribution of mass of holomorphic cusp forms. Duke Math. J. **162**(14), 2609–2644 (2013)
11. Blomer, V., Milićević, D.: Kloosterman sums in residue classes. J. Eur. Math. Soc. (JEMS) **17**(1), 51–69 (2015)
12. Choie, Y., Kohnen, W.: The first sign change of Fourier coefficients of cusp forms. Am. J. Math. **131**(2), 517–543 (2009)
13. Conrey, J.B., Farmer, D., Keating, J., Rubinstein, M., Snaith, N.: Integral moments of L-functions. Proc. Lond. Math. Soc. (3) **91**(1), 33–104 (2005)
14. Conrey, J.B., Iwaniec, H.: The cubic moment of central values of automorphic L-functions. Ann. Math. (2) **151**(3), 1175–1216 (2000)
15. Deligne, P.: Les constantes des équations fonctionelles des fonctions L. In: Modular Functions of One Variable, II (Proc. Internat. Summer School, University of Antwerp, Antwerp, 1972). Lecture Notes in Math. vol. 349. Springer, Berlin, pp. 501–597 (1973)
16. Duke, W.: Hyperbolic distribution problems and half-integral weight Maass forms. Invent. Math. **92**(1), 73–90 (1988)
17. Duke, W.: On ternary quadratic forms. J. Number Theory **110**(1), 37–43 (2005)
18. Duke, W., Schulze-Pillot, R.: Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids. Invent. Math. **99**(1), 49–57 (1990)
19. Folsom, A., Masri, R.: The asymptotic distribution of traces of Maass–Poincaré series. Adv. Math. **226**(4), 3724–3759 (2011)
20. Gallagher, P.X.: A large sieve density estimate near $\sigma = 1$. Invent. Math. **11**, 329–339 (1970)
21. Hoffstein, J., Kontorovich, A.: The first non-vanishing quadratic twist of an automorphic L-series (2010). arXiv:1008.0839 (**preprint**)
22. Hoffstein, J., Lockhart, P.: Coefficients of Maass forms and the Siegel zero. Ann. Math. (2) **140**(1), 161–181 (1994) (**With an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman**)
23. Iwaniec, H.: Small eigenvalues of Laplacian for $\Gamma_0(N)$. Acta Arith. **56**(1), 65–82 (1990)

24. Iwaniec, H.: Topics in Classical Automorphic Forms. Grad. Stud. Math., vol. 17. American Mathematical Society, Providence, RI (1997)
25. Iwaniec, H., Kowalski, E.: Analytic Number Theory. AMS Colloquium Publications, vol. 53. American Mathematical Society, Providence, RI (2004)
26. Iwaniec, H., Luo, W., Sarnak, P.: Low lying zeros of families of L-functions. Publ. Math. Inst. Hautes Études Sci. **91**, 55–131 (2001)
27. Kane, B.: Representations of integers by ternary quadratic forms. Int. J. Number Theory **6**(1), 127–159 (2010)
28. Kiral, E.M., Petrow, I., Young, M.: Oscillatory integrals with uniformity in parameters (2017). arXiv:1710.00916 (**preprint**)
29. Kiral, E.M., Young, M.: The fifth moment of modular L-functions (2017). arXiv:1701.07507 (**preprint**)
30. Kohnen, W.: Fourier coefficients of modular forms of half-integral weight. Math. Ann. **271**(2), 237–268 (1985)
31. Kowalski, E., Michel, P., VanderKam, J.: Mollification of the fourth moment of automorphic L-functions and arithmetic applications. Invent. Math. **142**(1), 95–151 (2000)
32. Li, W.: Newforms and functional equations. Math. Ann. **212**, 285–315 (1975)
33. Liu, S.-C., Masri, R., Young, M.: Subconvexity and equidistribution of Heegner points in the level aspect. Compos. Math. **149**, 1150–1174 (2013)
34. Liu, S.-C., Masri, R., Young, M.: Rankin–Selberg L-functions and the reduction of CM elliptic curves. Res. Math. Sci. **2**, 23 (2015) (**Art. 22**)
35. Masri, R.: Fourier coefficients of harmonic weak Mass forms and the partition function. Am. J. Math. **137**(4), 1061–1097 (2015)
36. Nelson, P.: Analytic isolation of newforms of given level. Arch. Math. (Basel) **108**(6), 555–568 (2017)
37. Ng, M.: The basis for space of cusp forms and Petersson trace formula. Masters thesis, University of Hong Kong (2012)
38. The on-line encyclopedia of integer sequences. published electronically at https://oeis.org (2016) (**sequence A008313**)
39. Pacetti, A.: On the change of root numbers under twisting and applications. Proc. Am. Math. Soc. **141**(8), 2615–2628 (2013)
40. Petrow, I.: A twisted Motohashi formula and Weyl-subconvexity for L-functions of weight two cusp forms. Math. Ann. **363**(1–2), 175–216 (2015)
41. Rouymi, D.: Formules de trace et non-annulation de fonctions L automorphes au niveau $\mathfrak{p}^{\nu}$. Acta Arith. **147**(1), 1–32 (2011)
42. The Sage Developers. SageMath, the Sage mathematics software system (2018). http://www.sagemath.org
43. Waldspurger, J.L.: Sur les coefficients de Fourier des formes modulaires de poids demi-entier. J. Math. Pures Appl. (9) **60**, 375–484 (1981)
44. Watson, G.N.: A Treatise on the Theory of Bessel Functions. Reprint of the second (1944) edition. Cambridge Mathematical Library, Cambridge University Press, Cambridge (1995)
45. Young, M.: Weyl-type hybrid subconvexity bounds for twisted L-functions and Heegner points on shrinking sets. J. Eur. Math. Soc. (JEMS) **19**(5), 1545–1576 (2017)
46. Young, M.: Explicit calculations with Eisenstein series (2017). arxiv:1710.03624 (**preprint**)