THE MOLLIFIED FOURTH MOMENT OF DIRICHLET L-FUNCTIONS

PENG GAO, XIAOSHENG WU, AND LIANGYI ZHAO

Abstract. We prove an asymptotic formula with a power saving error term for the fourth moment of the family of Dirichlet L-functions to modulus q mollified by a Dirichlet polynomial of length $q^{\frac{1}{22}-\varepsilon}$, valid for all moduli $q \not\equiv 2 \pmod 4$. This result was previously known only for restricted sets of moduli with smaller power savings. As a special case, when no Dirichlet polynomial is enclosed, this leads to a significant improvement on X. Wu's asymptotic evaluation of the fourth moment of Dirichlet L-functions at the cental point.

1. Introduction

It is an important subject to study moments of families of L-functions in number theory as they have a wide range of applications. In this paper, we focus on the classical setting of moments of the family of Dirichlet L-functions to a fixed modulus. Based on the work of J. B. Conrey, D. W. Farmer, J. P. Keating, M. O. Rubinstein and N. C. Snaith in [CFKRS05], it is conjectured that

$$M(k,q) := \sideset{}{^*}\sum_{\chi \pmod q} |L(\tfrac{1}{2},\chi)|^{2k} = \varphi^*(q)P_{k^2}(\log q) + O\left(q^{-\frac{1}{2}+\varepsilon}\right), \tag{1.1}$$

where $\sum^*$ indicates a sum over primitive Dirichlet characters χ modulo q, $L(s,\chi)$ the L-function attached to χ and $\varphi^*(q)$ the number of such primitive characters. We henceforth assume that $q \not\equiv 2 \pmod 4$ to ensure the existence of such primitive χ. Here, $P_{k^2}(x)$ is an explicitly computable polynomial

$$P_{k^2}(x) = a_{k^2}x^{k^2} + a_{k^2-1}x^{k^2-1} + \cdots + a_0$$

with coefficients in terms of finite Euler product

$$a_j = C_j \prod_{p|q}(1 + f_j(p)) \ll (\log\log q)^{O(1)}$$

for explicit constants C_j and arithmetic functions $f_j(p) \ll p^{-1}$. Throughout the paper, we reserve the letter p for a prime number. The explicit expressions of a_j can be calculated from the shifted moments by setting all shifts to zero (c.f Theorem 1.1 for the case $k = 2$). In particular, when q is a prime, all coefficients a_j are absolute constants. In particular, (1.1) implies that asymptotically,

$$M(k,q) \sim a_{k^2}\varphi^*(q)(\log q)^{k^2}, \tag{1.2}$$

which is also expected (see [RS05]) to hold for all real $k \geq 0$.

2010 *Mathematics Subject Classification.* 11M06, 11F72 .

Key words and phrases. Mollified fourth moment; Kloosterman sum; Dirichlet L-functions; Critical line; Divisor problem.

The validity of (1.2) for $k = 1$ follows from a result of A. Selberg [Sel46] on a more general twisted second moment. Asymptotic evaluation (1.2) for $k = 2$ was done by D. R. Heath-Brown [HB81] for most all q, whose result was subsequently extended to all q by K. Soundararajan [Sou07]. For prime moduli q, M. P. Young [You11] was the first to obtain the asymptotic formula (1.1) for $M(k, q)$ when $k = 2$ with a power saving error term $q^{-\frac{5}{512}+\varepsilon}$. The error term in Young's result are subsequently improved in [BFKMM17+a, BFKMM17+b, BFKMMS23] to $q^{-\frac{1}{20}+\varepsilon}$, which is currently the best known result for prime moduli. For general moduli q, the asymptotic formula (1.1) was later obtained by X. Wu [Wu23+a] with a power saving error term $q^{-\frac{11}{448}+\varepsilon}$.

Instead of studying $M(k, q)$, one may consider more generally the so-called twisted shifted moments. For example, we say the corresponding case for $k = 2$ is the twisted shifted fourth moment, which has the form

$$\sum_{\chi \,(\mathrm{mod}\ q)}^{*} L\left(\tfrac{1}{2} + \alpha, \chi\right) L\left(\tfrac{1}{2} + \beta, \chi\right) L\left(\tfrac{1}{2} + \gamma, \overline{\chi}\right) L\left(\tfrac{1}{2} + \delta, \overline{\chi}\right) \chi(h)\overline{\chi}(k), \tag{1.3}$$

where α, β, γ and δ are complex numbers, and $h, k \in \mathbb{Z}$.

For prime moduli q, the expression in (1.3) was asymptotically evaluated for square-free h, k by B. Hough [Hou16, Theorem 4]. The result was extended to cube-free h, k by R. Zacharias [Zac19], and to general h, k by D. Liu [Liu24, Theorem 6.2]. In [GZ25+b], P. Gao and L. Zhao obtained an asymptotical formula for the expression in (1.3) for q being certain prime powers and general h, k.

Evaluation of the twisted shifted moments is essential towards our understanding of the so-called mollified moments, which, for $k = 2$, has the form

$$\sum_{\chi \,(\mathrm{mod}\ q)}^{*} L\left(\tfrac{1}{2} + \alpha, \chi\right) L\left(\tfrac{1}{2} + \beta, \chi\right) L\left(\tfrac{1}{2} + \gamma, \overline{\chi}\right) L\left(\tfrac{1}{2} + \delta, \overline{\chi}\right) \left|A\left(\tfrac{1}{2}, \chi\right)\right|^2, \tag{1.4}$$

where $A\left(\frac{1}{2}, \chi\right)$ is a Dirichlet polynomial

$$A\left(\tfrac{1}{2}, \chi\right) = \sum_{h \le y} \frac{\alpha_h \chi(h)}{\sqrt{h}} \tag{1.5}$$

with the coefficients satisfying $\alpha_h \ll h^{\varepsilon}$.

The mollified moment plays a pivotal role in an extensive range of problems, and asymptotic formulas with a power saving error term are essential prerequisites for many applications including the techniques of amplification, modification or resonance. For example, they have been applied to study extreme values of L-functions [Sou08], non-vanishing of L-functions at the central point [CS07], and to achieve sharp bounds for moments of L-functions with real exponents [RS15].

The degenerate case when $h = k = 1$ in (1.3) is called the shifted fourth moment. One may define shifted moments for other families of L-functions as well. As revealed in [Mun17, Sza24], these shifted moments are especially useful in estimating moments of the corresponding character sums.

In view of the importance of the twisted shifted moments and mollified moments, it is the aim of this paper to evaluate asymptotically the expressions in (1.3) and (1.4) for general moduli q. As the situations are similar, we consider only the case of even characters (i.e. those characters χ with $\chi(-1) = 1$) throughout the paper. For this, we define

$$\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta) := \frac{2}{\varphi^*(q)} \sideset{}{^+}\sum_{\chi \,(\mathrm{mod}\ q)} L\left(\tfrac{1}{2}+\alpha,\chi\right) L\left(\tfrac{1}{2}+\beta,\chi\right) L\left(\tfrac{1}{2}+\gamma,\overline{\chi}\right) L\left(\tfrac{1}{2}+\delta,\overline{\chi}\right) \chi(h\overline{k}),$$

$$\mathcal{M}(\alpha,\beta,\gamma,\delta) := \frac{2}{\varphi^*(q)} \sideset{}{^+}\sum_{\chi \,(\mathrm{mod}\ q)} L\left(\tfrac{1}{2}+\alpha,\chi\right) L\left(\tfrac{1}{2}+\beta,\chi\right) L\left(\tfrac{1}{2}+\gamma,\overline{\chi}\right) L\left(\tfrac{1}{2}+\delta,\overline{\chi}\right) \left|A\left(\tfrac{1}{2},\chi\right)\right|^2,$$

where $\overline{k}$ denotes the inverse of k modulo q, the symbol $\sum^+$ henceforth indicates that the summation is over all primitive even characters, and A is defined in (1.5).

To state our result, we need to introduce some notations. Let

$$X_{\alpha,\gamma} = X_\alpha X_\gamma, \qquad X_{\alpha,\beta,\gamma,\delta} = X_\alpha X_\beta X_\gamma X_\delta$$

with

$$X_\alpha = \left(\frac{q}{\pi}\right)^{-\alpha} \frac{\Gamma\left(\frac{\frac{1}{2}-\alpha}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\alpha}{2}\right)}.$$

Set

$$Z_{h,k,q}(\alpha,\beta,\gamma,\delta) = Y_h(\alpha,\beta,\gamma,\delta) Y_k(\gamma,\delta,\alpha,\beta) Z_q(\alpha,\beta,\gamma,\delta), \tag{1.6}$$

where

$$Z_q(\alpha,\beta,\gamma,\delta) = \frac{\zeta_q(1+\alpha+\gamma)\zeta_q(1+\alpha+\delta)\zeta_q(1+\beta+\gamma)\zeta_q(1+\beta+\delta)}{\zeta_q(2+\alpha+\beta+\gamma+\delta)}$$

with $\zeta_q(s)$ the Euler product of $\zeta(s)$ with the primes dividing q removed and

$$Y_a(\alpha,\beta,\gamma,\delta) = \frac{1}{a^\gamma} \sum_{d|a} d^{\gamma-\delta} \prod_{p|d} \left(1 - \frac{1}{p^{1+\alpha+\gamma}}\right)\left(1 - \frac{1}{p^{1+\beta+\gamma}}\right)\left(1 - \frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right)^{-1}. \tag{1.7}$$

It is easy to see that $Z_q(\alpha,\beta,\gamma,\delta)$ is invariant under transpositions $\alpha \leftrightarrow \beta$ and $\gamma \leftrightarrow \delta$, and so is $Y_a(\alpha,\beta,\gamma,\delta)$ (see Lemma 4.1 below). As a result, this invariance under transpositions $\alpha \leftrightarrow \beta$ and $\gamma \leftrightarrow \delta$ also hold for $Z_{h,k,q}(\alpha,\beta,\gamma,\delta)$.

1.1. **Main results.**

Now, we are ready to state our results on $\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta)$ and $\mathcal{M}(\alpha,\beta,\gamma,\delta)$. We begin with $\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta)$.

Theorem 1.1. *With the notation as above, let h, k be integers satisfying $(h,k) = (hk,q) = 1$. Then there exists $\eta > 0$ such that for $\alpha,\beta,\gamma,\delta \in \{z \in \mathbb{C} : \Re(z) < \eta/\log q\}$, we have*

$$\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta) = (hk)^{-\frac{1}{2}} \mathfrak{M}_{h,k}(\alpha,\beta,\gamma,\delta) + O\left(q^{-\frac{1}{20}+\varepsilon}(|h|+|k|)^{\frac{3}{10}} \Delta^{O(1)}\right), \tag{1.8}$$

where

$$\begin{aligned}\Delta =&(1+|\alpha|)(1+|\beta|)(1+|\gamma|)(1+|\delta|), \text{ and}\\ \mathfrak{M}_{h,k}(\alpha,\beta,\gamma,\delta) =& Z_{h,k,q}(\alpha,\beta,\gamma,\delta)+X_{\alpha,\beta,\gamma,\delta}Z_{h,k,q}(-\gamma,-\delta,-\alpha,-\beta)\\ &+X_{\alpha,\gamma}Z_{h,k,q}(\beta,-\gamma,\delta,-\alpha)+X_{\beta,\gamma}Z_{h,k,q}(\alpha,-\gamma,\delta,-\beta)\\ &+X_{\alpha,\delta}Z_{h,k,q}(\beta,-\delta,\gamma,-\alpha)+X_{\beta,\delta}Z_{h,k,q}(\alpha,-\delta,\gamma,-\beta).\end{aligned}$$

Our next result evaluates $\mathcal{M}(\alpha,\beta,\gamma,\delta)$.

Theorem 1.2. *With the notation as above, there exists $\eta > 0$ such that for $\alpha,\beta,\gamma,\delta \in \{z \in \mathbb{C} : \Re(z) < \eta/\log q\}$, we have*

$$\mathcal{M}(\alpha,\beta,\gamma,\delta) = \sum_{\substack{ah,ak\le y\\ (ahk,q)=1\\ (h,k)=1}} \frac{\alpha_{ah}\overline{\alpha_{ak}}}{ahk}\mathfrak{M}_{h,k}(\alpha,\beta,\gamma,\delta) + O\left(q^{-\frac{1}{20}+\varepsilon}y^{\frac{11}{10}}\Delta^{O(1)}\right).$$

Remark. To facilitate the reader's understanding of the strength of Theorems 1.1–1.2, we make some remarks here.

- If one sets $h = k = 1$ and all shifts to zero, then Theorem 1.1 establishes an asymptotic formula for the fourth moment of Dirichlet L-functions as in [Wu23+a], but with the error term being sharpened to $q^{-\frac{1}{20}+\varepsilon}$. This error term is as strong as the best known result established for the prime moduli case in [BFKMM17+a, BFKMM17+b, BFKMMS23]. Given that our method of treating the error terms is quite different from the those carried out in [BFKMM17+a, BFKMM17+b, BFKMMS23] concerning the prime moduli case, curiosity naturally arises regarding the reason for this identical end via vastly distinct means. We note here that upon careful comparison, the term $q^{-\frac{1}{20}+\varepsilon}$ in both situations ultimately stems from certain diagonal terms introduced when applying the Cauchy-Schwarz inequality. This suggests, to a certain extent, that any further improvement in the q-aspect seems very unlikely via these approaches.
- For twisted moments, we contrast Theorem 1.1 with previously established results for various special cases, such as [Hou16, Zac19, Liu24, GZ25+b]. We note that Theorem 1.1 establishes the asymptotic formula for general moduli q and general h, k, sharping the error term in all aspects including q, h, and k.
- The asymptotic formula of the mollified moment in Theorem 1.2 is valid (the main term dominates the O-term) if the length of the mollifier does not exceed $q^{\frac{1}{22}-\varepsilon}$. This situation is somewhat analogous to the mollified fourth moment of the Riemann zeta-function, established in the works of C. P. Hughes and P. Y. Matthew [HM10] and S. Bettin, H. M. Bui, X. Li, and M. Radziwiłł [BBLR16], where stationary phase methods applied to the Archimedean integral over t impose stringent constraints on the shifted convolution sum through the variable clustering effect, thereby producing significantly improved error terms.
- If one only aims for an upper bound, then we note that V. Blomer, P. Humphries, R. Khan, and M. B. Milinovich [BHKM20] applied their Motohashi's fourth moment identity to obtain that

$$M(0,0,0,0) \ll_\varepsilon q^\varepsilon,$$

for prime moduli q and square-free sequences $\alpha_h \ll h^{\varepsilon}$. Their upper bound holds for a very long mollifier with $y \ll q^{\frac{1}{4}-\varepsilon}$. However, as mentioned in [BHKM20], their method does not give asymptotic results.

- In view of Theorem 1.1, Theorem 1.2 gains some additional power saving in the y-aspect through averaging over h and k.
- The utility and relevance of Theorems 1.1–1.2 clearly lie in the study of the family of Dirichlet L-functions (more specifically, non-vanishing at $s = 1/2$ and existence of large values, for instance). As a direct application, one may apply the approach in the work of Gao and Zhao [GZ25+a], while using the result of Theorem 1.2 in place of those in [Hou16] or [Zac19]. This allows one to extend the result [GZ25+a] to establish sharp upper bounds for all $2k$-th moment for $k \leq 2$ at the central point of the family of Dirichlet L-functions to general moduli.

Our proofs of Theorems 1.1–1.2 loosely follow the *modus operandi* of [You11] and [Wu23+a]. The calculation of the main terms is quite involved, and their discernment, most crucially, comes from studying a smoothed version of the generalized Esterman D-functions (discussed in Section 2.4). The treatments for the error terms are divided into two parts: the balanced ones and the unbalanced ones. The balanced error terms are handled using spectral theory and the spectral large sieve, where we remove the dependence of the Ramanujan–Petersson conjecture and minimize the impact from h, k by establishing new upper bounds for sums involving Kloosterman sums, Theorems 1.3–1.4. The treatment of the unbalanced error terms is based on applying new bound for bilinear forms of incompleted Kloosterman sums, which was proved for general moduli by B. Kerr, I. E. Shparlinski, X. Wu, and P. Xi [KSWX23].

We note that as our primary goal is to obtain power saving error terms in the q aspect when evaluating $\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta)$ and $\mathcal{M}(\alpha,\beta,\gamma,\delta)$. Thus we do not make explicit the dependence on the parameters $\alpha,\beta,\gamma,\delta$ in the error terms. Often, in our treatments below, we shall omit the dependency on these parameters as well.

1.2. **Upper bounds for sums of Kloosterman sums.** In our proof of Theorems 1.1–1.2, we need to estimate certain shifted convolution sums that appear in the process (see Section 8), which requires a good control on related sums involving with Kloosterman sums. We develop the necessary tools in this paper as well.

Let $\mathfrak{a}$ and $\mathfrak{b}$ denote two cusps of the Hecke congruence group $\Gamma_0(Q)$, with corresponding scaling matrixes $\sigma_{\mathfrak{a}}, \sigma_{\mathfrak{b}}$. For integers m, n and real number γ such that there exists a matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma_0(Q)\sigma_{\mathfrak{b}}$, the associated Kloosterman sum $S_{\mathfrak{a},\mathfrak{b}}(m,n;\gamma)$ is defined as

$$S_{\mathfrak{a},\mathfrak{b}}(m,n;\gamma) = \sideset{}{'}\sum_{\delta \,(\mathrm{mod}\ \gamma\mathbb{Z})} e\left(m\frac{\alpha}{\gamma} + n\frac{\delta}{\gamma}\right), \tag{1.9}$$

where the sum runs over all δ modulo $\gamma\mathbb{Z}$ for which there exist α, β completing the matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $\sigma_{\mathfrak{a}}^{-1}\Gamma_0(Q)\sigma_{\mathfrak{b}}$. When $\mathfrak{a} = \mathfrak{b} = \infty$ and $Q = q$, this reduces to the classical Kloosterman sum $S(m,n;q)$.

Sums of the type

$$\sum_{M<m\le 2M} a_m \sum_{N<n\le 2N} b_n \sum_{\gamma} \frac{1}{\gamma}\phi\left(\frac{4\pi\sqrt{mn}}{\gamma}\right) S_{\mathfrak{a},\mathfrak{b}}(m,\pm n;\gamma) \tag{1.10}$$

have been widely studied due to their numerous applications. Here ϕ is a smooth test function satisfying

$$\mathrm{Supp}\phi \subset [X, 8X], \tag{1.11}$$

and

$$\|\phi\|_\infty \ll 1, \quad \|\phi'\|_1 \ll 1, \quad \|\phi''\|_1 \ll X^{-1}, \tag{1.12}$$

for some positive parameter X. The classical upper bound for such sums was established by J. M. Deshouillers and H. Iwaniec [DI82, Theorem 8], using spectral analysis.

The asymptotic twisted moment problem requires evaluating the shifted convolution sum

$$\sum\sum_{\substack{hm\equiv kn \,(\mathrm{mod}\ q)\\ (mn,q)=1,\ hm\neq kn}} \tau(m)\tau(n)W\left(\frac{m}{M}\right)W\left(\frac{n}{N}\right),$$

for $h,k,M,N \ge 1$ and smooth test functions W. The error terms in such evaluation depend critically on estimating sums of the form (1.10) involving Kloosterman sums $S_{\infty,\mathfrak{b}}(sm,\pm n;\gamma)$ with an individual shift s (typically large). Here, the level of the Hecke congruence group Q is roughly a product of hk with certain factors of q arising from the comprimality condition $(mn,q) = 1$. Usually, the spectral analysis of such Kloosterman sums with an individual shift will produce error terms involving s-power factors, whose exponents depend on the current state of our knowledge towards the Ramanujan–Petersson conjecture. When establishing an asymptotic for the second moment of twisted modular L-functions, one must handle a similar shifted convolution sum

$$\sum\sum_{\substack{m\equiv n \,(\mathrm{mod}\ q)\\ (mn,q)=1,\ m\neq n}} \lambda_f(m)\lambda_g(n)W\left(\frac{m}{M}\right)W\left(\frac{n}{N}\right),$$

where f and g denote fixed cuspidal Hecke eigenforms, with respective Hecke eigenvalues λ_f and λ_g. For this sum, Blomer and Milićević [BM15] successfully eliminated the dependence on the Ramanujan–Petersson conjecture through a clever arrangement of Hölder's inequality after the spectral decomposition. We are able to adapt their idea to our present setting, incorporating new techniques to manage the higher level Q's. The key input is Theorem 11.6, where we minimize the influence of s through a novel relation of Fourier coefficients, Proposition 11.4.

Let $\lambda_1(Q)$ denote the least positive eigenvalue of the Laplacian for $\Gamma_0(Q)$, and we define the number

$$\vartheta_Q = \sqrt{\max\{0, 1-4\lambda_1(Q)\}}.$$

Selberg's eigenvalue conjecture posits that $\lambda_1(Q) \ge \frac{1}{4}$, which directly implies $\vartheta_Q = 0$. To date, the most comprehensive understanding we have regarding ϑ_Q in general is that $\vartheta_Q \le \frac{1}{2}$. Our results on Kloosterman sums are presented in the following two theorems. These results can be of independent interest since they may be applied not only to analogous shifted convolution sums but also to other problems in analytic number theory.

Theorem 1.3. *We keep the notation as above. Let $s \in \mathbb{N}$, $Q = uv$ with $(u, v) = 1$, $M, N \geq 1$, and set $C = M + N + Q + X + s$. Suppose that ϕ is a real valued function satisfying (1.11) and (1.12). For any complex sequence $(a_m)_{m\in[M,2M]}, (b_n)_{n\in[N,2N]}$, we have*

(1.13)
$$\sum_m a_m \sum_n b_n \sum_\gamma \frac{1}{\gamma}\phi\left(\frac{4\pi\sqrt{smn}}{\gamma}\right) S_{\infty,1/u}(sm, \pm n; \gamma) \ll C^\varepsilon \frac{1 + |\log X| + X^{-\vartheta_Q}}{1 + X}$$
$$\times\left(1 + X + \frac{s^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1 + X + \frac{(s, Q)^{\frac{1}{2}} M}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1 + X + \frac{N^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right) M^{\frac{1}{2}} \|a_m\|_\infty^{\frac{1}{2}} \|b_n\|_2.$$

Through standard Fourier analysis (for instance, as in the proof for [DI82, Theorem 13, p.269]), one can derive from Theorem 1.3 the subsequent result.

Theorem 1.4. *Keeping the notation as in Theorem 1.3, suppose that $g(m, n, l)$ is a real valued function of $\mathscr{C}^6$ class, having compact support in $[M, 2M] \times [N, 2N] \times [L, 2L]$ and satisfying*

$$l^{j_1} m^{j_2} n^{j_3} \frac{\partial^{j_1+j_2+j_3}}{\partial l^{j_1} \partial m^{j_2} \partial n^{j_3}} g(m, n, l) \ll C^\varepsilon \quad \text{with } 0 \leq j_1, j_2, j_3 \leq 2.$$

For any complex sequences a_m, b_n, we have

$$\sum_{(l,v)=1} \sum_m a_m \sum_n b_n g(m, n, l) S(sm\bar{v}, \pm n; lu) \ll C^\varepsilon \frac{1 + |\log X| + X^{-\vartheta_Q}}{1 + X} uv^{\frac{1}{2}} L$$
$$\times\left(1 + X + \frac{s^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1 + X + \frac{(s, Q)^{\frac{1}{2}} M}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1 + X + \frac{N^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right) M^{\frac{1}{2}} \|a_m\|_\infty^{\frac{1}{2}} \|b_n\|_2,$$

where

$$X = \frac{(sMN)^{\frac{1}{2}}}{uv^{\frac{1}{2}} L}.$$

1.3. **Notation.** As usual, we use ε to denote an arbitrarily small positive constant that may vary from place to place. We write $W(x)$ for a smooth non-negative function, which may have different expressions for each occurrence, but always has compact support in $[1/2, 2]$ and satisfies

(1.14)
$$W^{(j)}(x) \ll q^{j\varepsilon}$$

for any $j \geq 0$. We denote its Mellin transform by $\widetilde{W}$ and its Fourier transform by $\widehat{W}$, with similar notation for other functions. For $d, q \in \mathbb{N}$, denote by q_d the maximal factor of q such that $(q_d, d) = 1$.

The notations $\sigma_\lambda(n)$ and $\sigma_{\alpha,\beta}(n)$ are defined as

$$\sigma_\lambda(n) = \sum_{d|n} d^\lambda, \quad \sigma_{\alpha,\beta}(n) = \sum_{ad=n} a^{-\alpha} d^{-\beta}.$$

Clearly, $\sigma_{\alpha,\beta}(n) = n^{-\alpha}\sigma_{\alpha-\beta}(n)$. We also write $\tau(n)$ for the divisor function.

2. Auxiliary lemmas

2.1. **Approximate functional equation.** We shall evaluate $M_{h,k}(\alpha,\beta,\gamma,\delta)$ by first expanding it into a Dirichlet L-series, using the following approximate functional equation, derived from the standard functional equation of $L(s,\chi)$. This result is taken from [You11, Proposition 2.4].

Lemma 2.1 (Approximate functional equation)**.** *Let $G(s)$ be an even, entire function of exponential decay in any strip $|Re(s)| < C$ satisfying*

$$G(0) = 1, \quad G\left(\tfrac{1}{2} \pm \alpha\right) = G\left(\tfrac{1}{2} \pm \beta\right) = G\left(\tfrac{1}{2} \pm \gamma\right) = G\left(\tfrac{1}{2} \pm \delta\right) = 0.$$

Define

$$V_{\alpha,\beta,\gamma,\delta}(x) = \frac{1}{2\pi i}\int_{(1)} \frac{G(s)}{s} g_{\alpha,\beta,\gamma,\delta}(s) x^{-s} ds,$$

where

$$g_{\alpha,\beta,\gamma,\delta}(s) = \pi^{-2s} \frac{\Gamma\left(\frac{\frac{1}{2}+\alpha+s+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\beta+s+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\gamma+s+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\delta+s+\mathfrak{a}}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\alpha+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\beta+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\gamma+\mathfrak{a}}{2}\right)\Gamma\left(\frac{\frac{1}{2}+\delta+\mathfrak{a}}{2}\right)}, \quad \mathfrak{a} = \frac{1-\chi(-1)}{2}.$$

We have

$$\begin{aligned} L\left(\tfrac{1}{2}+\alpha,\chi\right)&L\left(\tfrac{1}{2}+\beta,\chi\right)L\left(\tfrac{1}{2}+\gamma,\overline{\chi}\right)L\left(\tfrac{1}{2}+\delta,\overline{\chi}\right) \\ &= \sum_{m,n} \frac{\sigma_{\alpha,\beta}(m)\sigma_{\gamma,\delta}(n)\chi(m)\overline{\chi}(n)}{(mn)^{\frac{1}{2}}} V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn}{q^2}\right) \\ &\qquad + X_{\alpha,\beta,\gamma,\delta}\sum_{m,n} \frac{\sigma_{-\alpha,-\beta}(m)\sigma_{-\gamma,-\delta}(v)\overline{\chi}(m)\chi(n)}{(mn)^{\frac{1}{2}}} V_{-\alpha,-\beta,-\gamma,-\delta}\left(\frac{mn}{q^2}\right). \end{aligned}$$

2.2. **The orthogonality formula.** After using the approximate functional equation to expand $M_{h,k}(\alpha,\beta,\gamma,\delta)$, we need to compute various sums over characters via a standard orthogonality formula presented below, available in [HB81] and [Sou07].

Lemma 2.2 (The orthogonality formula)**.** *For $(mn,q) = 1$, we have*

$$\sideset{}{^+}\sum_{\chi(\bmod q)} \chi(m)\overline{\chi}(n) = \tfrac{1}{2}\sum_{d|(q,m\pm n)} \varphi(d)\mu\left(\frac{q}{d}\right).$$

2.3. **Two partitions of unity.** We also want to localize the variables in certain sums appearing in the proof of Theorems 1.1–1.2. To that end, we utilize two smooth partitions of unity. The first is the classical dyadic partition.

Lemma 2.3 (Smooth dyadic partition of unity)**.** *There exists a smooth non-negative function $W(x)$ having a support on* $[1/2,2]$ *and satisfying* (1.14) *such that*

$$\sum_{k\geq 0} W\left(\frac{x}{2^k}\right) = 1, \tag{2.1}$$

for any $x \geq 1$.

Our next lemma introduces a smooth partition designed to distinguish the relative sizes of two variables. This partition is also discussed in [BBLR16] and [Wu23+b, Section 2.6].

Lemma 2.4. *There is a smooth function $\omega(x)$ such that*

$$\omega(x) + \omega\left(x^{-1}\right) = 1,$$

for any $x \in \mathbb{R}$, and $\omega(x) \ll_j (1+x)^{-j}$ for any fixed $j > 0$ and $x > 1$. In addition, its Mellin transform $\widetilde{\omega}(u)$ has a simple pole at $u = 0$ with residue 1, and satisfies

$$\widetilde{\omega}\left(\pm\frac{\alpha-\beta}{2}\right) = \widetilde{\omega}\left(\pm\frac{\gamma-\delta}{2}\right) = 0.$$

2.4. **Generalized Esterman D-function and Voronoi summation formula.** Let $V(x)$ be a bounded function. For $q, l \in \mathbb{N}$, $h \in \mathbb{Z}$, and $\lambda \in \mathbb{C}$, define the generalized Esterman D-function as

$$D_q\left(s, \lambda, \frac{h}{l}; V\right) = \sum_{(n,q)=1} \frac{\sigma_\lambda(n)}{n^s} e\left(n\frac{h}{l}\right) V(n). \tag{2.2}$$

We also adopt the convention by writing $D_q\left(s, \lambda, \frac{h}{l}\right)$ for $D_q\left(s, \lambda, \frac{h}{l}; V\right)$ when $V(x) \equiv 1$. This function was originally investigated in [Wu19] and further extended in [Wu23+a].

Lemma 2.5. *Let q, h, l be integers, satisfying $(l, hq) = 1$. For any fixed $\lambda \in \mathbb{C}$, $D_q\left(s, \lambda, \frac{h}{l}\right)$ is meromorphic as a function of s, satisfying the functional equation*

$$\begin{aligned} D_q\left(s, \lambda, \frac{h}{l}\right) = {} & 2(2\pi)^{-2-\lambda+2s} \Gamma(1-s)\Gamma(1+\lambda-s) \frac{\varphi(q)}{q} \sum_{a|q} \frac{\mu^2(a)}{\varphi(a)} (la)^{1+\lambda-2s} \\ & \times \left[\cos\left(\frac{\pi\lambda}{2}\right) D_a\left(1-s, -\lambda, \frac{\overline{ha^2}}{l}\right) - \cos\left(\pi\left(s-\frac{\lambda}{2}\right)\right) D_a\left(1-s, -\lambda, -\frac{\overline{ha^2}}{l}\right)\right]. \end{aligned} \tag{2.3}$$

If $\lambda \neq 0$, then $D_q(s, \lambda, \frac{h}{l})$ has simple poles at $s = 1$ and $s = 1+\lambda$ with the corresponding residues being, respectively,

$$\frac{\varphi(q)}{q} l^{-1+\lambda} \zeta_q(1-\lambda), \qquad \frac{\varphi(q)}{q} l^{-1-\lambda} \zeta_q(1+\lambda).$$

Proof. This arises as the case $r = 0$ of [Wu23+a, Proposition 5.3], with residues being immediate. In the functional equation given there, the b-sums all equal 1 for our case. Upon expanding the Esterman D-function as a Dirichlet series, we factor the exponential

$$e\left(n\frac{\overline{h_i}}{la_1}\right) = e\left(n\frac{\overline{ha_1^2}}{l}\right) e\left(n\frac{\overline{il^2}}{a_1}\right).$$

The i-sum evaluates to $\mu(a_1)$ via Ramanujan summation. For fixed a_1, the a-sum in [Wu23+a, Proposition 5.3] becomes

$$\sum_{a_1|a|q} \frac{\mu(a)}{a} = \frac{\varphi(q)}{q} \frac{\mu(a_1)}{\varphi(a_1)}.$$

Substituting the above into the original functional equation given in [Wu23+a, Proposition 5.3] directly yields (2.3). This completes the proof. □

Lemma 2.6 (Voronoi formula)**.** *Let q, h, l be integers satisfying $(l, hq) = 1$. For a smooth, compactly supported function $\mathcal{V} : (0, \infty) \to \mathbb{C}$ and any complex number $\lambda \neq 0$, we have*

$$
\begin{aligned}
(2.4)\quad D_q\left(s, \lambda, \frac{h}{l}; \mathcal{V}\right) = &\frac{\varphi(q)}{q} l^{-1+\lambda} \zeta_q(1-\lambda) \widetilde{\mathcal{V}}(1-s) + \frac{\varphi(q)}{q} l^{-1-\lambda} \zeta_q(1+\lambda) \widetilde{\mathcal{V}}(1-s+\lambda) \\
&+ \frac{\varphi(q)}{q} \sum_{a|q} \frac{\mu^2(a)}{\varphi(a)} (la)^{-1+\lambda} \sum_{(m,a)=1} \sigma_{-\lambda}(m) e\left(\pm m \frac{\overline{ha^2}}{l}\right) \mathring{\mathcal{V}}_{\pm}\left(\frac{m}{(la)^2}\right),
\end{aligned}
$$

where

$$
\mathring{\mathcal{V}}_{\pm}(x) = 2(2\pi)^{-2-\lambda} \frac{1}{2\pi i} \int_{(-2)} (4\pi^2 x)^{u-1} \widetilde{\mathcal{V}}(u-s) \Gamma(1-u) \Gamma(1+\lambda-u) S_{\pm} \mathrm{d}u,
$$

with

$$
S_+ = \cos\left(\frac{\pi\lambda}{2}\right), \quad S_- = \cos\left(\pi\left(u - \frac{\lambda}{2}\right)\right).
$$

If $\mathcal{V}(x) = W(x/M)$, then

$$
x^j \mathring{\mathcal{V}}_{\pm}^{(j)}(x) \ll_{A,j} q^{j\varepsilon} (1 + Mx)^{-A} M^{Re(s)}.
$$

Proof. Applying the Mellin transform of $\mathcal{V}$, we obtain

$$
\begin{aligned}
D_q\left(s, \lambda, \frac{h}{l}; \mathcal{V}\right) &= \frac{1}{2\pi i} \int_{(2)} \widetilde{\mathcal{V}}(u) D_q\left(s+u, \lambda, \frac{h}{l}\right) \mathrm{d}u \\
&= \frac{1}{2\pi i} \int_{(2)} \widetilde{\mathcal{V}}(u-s) D_q\left(u, \lambda, \frac{h}{l}\right) \mathrm{d}u.
\end{aligned}
$$

Shifting the contour to $c_u = -2$, the residues at $u = 1$ and $u = 1 + \lambda$ (given by Lemma 2.5) yield

$$
\frac{\varphi(q)}{q} l^{-1+\lambda} \zeta_q(1-\lambda) \widetilde{\mathcal{V}}(1-s), \quad \frac{\varphi(q)}{q} l^{-1-\lambda} \zeta_q(1+\lambda) \widetilde{\mathcal{V}}(1-s+\lambda),
$$

respectively. The remaining integral is handled by applying the functional equation in Lemma 2.5, and then expanding the resulting D_q using the Dirichlet series expression given in (2.2), which leads to (2.4) .

When $\mathcal{V}(x) = W(x/M)$, we have

$$
\widetilde{\mathcal{V}}(u) = M^u \widetilde{W}(u),
$$

and then

$$
(2.5)\quad \mathring{\mathcal{V}}_{\pm}(x) = 2(2\pi)^{-2-\lambda} M^s \frac{1}{2\pi i} \int_{(-2)} (4\pi^2 M x)^{u-1} \widetilde{W}(u-s) \Gamma(1-u) \Gamma(1+\lambda-u) S_{\pm} \mathrm{d}u.
$$

The bound

$$
\mathring{\mathcal{V}}_{\pm}(x) \ll_A q^{\varepsilon} (1 + Mx)^{-A} M^{\mathrm{Re}(s)}
$$

follows from shifting the line of integration in (2.5) to $c_u = -A$ ($Mx \gg q^{\varepsilon}$) and $c_u = 1 - \varepsilon$ ($Mx \ll q^{\varepsilon}$). The bounds on the derivatives follow by differentiating j times under the integral sign. This completes the proof. □

3. Sketch of the treatment for moments

Upon applying the approximate functional equation (Lemma 2.1) and then evaluating the resulting character sums via orthogonality (Lemma 2.2), we see that $\mathcal{M}_{h,k}$ is decomposed as

$$\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta) = \mathscr{M}_{h,k}(\alpha,\beta,\gamma,\delta) + \mathscr{M}_{\overline{h,k}}(\alpha,\beta,\gamma,\delta), \tag{3.1}$$

where

$$\mathscr{M}_{h,k}(\alpha,\beta,\gamma,\delta) = \frac{1}{\varphi^*(q)} \sum_{d|q} \varphi(d)\mu\left(\frac{q}{d}\right) \times \sum_{\substack{hm\equiv\pm kn \ (\mathrm{mod}\ d)\\ (mn,q)=1}} \frac{1}{m_1^{\frac12+\alpha} m_2^{\frac12+\beta} n_1^{\frac12+\gamma} n_2^{\frac12+\delta}} V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn}{q^2}\right),$$

where we set $m = m_1m_2$, $n = n_1n_2$, and

$$\mathscr{M}_{\overline{h,k}}(\alpha,\beta,\gamma,\delta) = X_{\alpha,\beta,\gamma,\delta}\mathscr{M}_{k,h}(-\alpha,-\beta,-\gamma,-\delta). \tag{3.2}$$

We further split $\mathscr{M}_{h,k}$ as

$$\mathscr{M}_{h,k} = \mathscr{M}^D_{h,k} + \mathscr{M}^+_{h,k} + \mathscr{M}^-_{h,k}, \tag{3.3}$$

where $\mathscr{M}^D_{h,k}$ corresponds to the contribution from the diagonal terms ($hm = kn$), while $\mathscr{M}^+_{h,k}$ and $\mathscr{M}^-_{h,k}$ account for the contribution from off-diagonal terms with $hm \equiv kn \pmod d$ and $hm \equiv -kn \pmod d$ respectively. The term $\mathscr{M}^D_{h,k}$ contributes to a main term and will be treated in Section 5.

Similarly, we have

$$\mathscr{M}_{\overline{h,k}} = \mathscr{M}^D_{\overline{h,k}} + \mathscr{M}^+_{\overline{h,k}} + \mathscr{M}^-_{\overline{h,k}}.$$

In what follows, we shall primarily focus on the evaluation of $\mathscr{M}_{h,k}$ as $\mathscr{M}_{\overline{h,k}}$ can be evaluated via the relation (3.2).

Upon applying Möbius inversion to eliminate the conditions $(m_i, q_d) = 1$ and $(n_i, q_d) = 1$, we recast the two non-diagonal terms as

$$\mathscr{M}^\pm_{h,k}(\alpha,\beta,\gamma,\delta) = \frac{1}{\varphi^*(q)} \sum_{d|q} \varphi(d)\mu\left(\frac{q}{d}\right) \sum_{\substack{abc|q_d^2\\ (a,b)=1}} \varpi_{\alpha,\beta}(ac,q_d)\varpi_{\gamma,\delta}(bc,q_d)A^\pm_{ah,bk}\left(d,\mathcal{X}\right), \tag{3.4}$$

where

$$\varpi_{\alpha,\beta}(a,q) = \sum_{\substack{a_1|q,\ a_2|q\\ a_1a_2=a}} \frac{\mu(a_1)\mu(a_2)}{a_1^{\frac12+\alpha} a_2^{\frac12+\beta}} \ll a^{-\frac12+\varepsilon}, \quad \mathcal{X} = \frac{q^2}{abc^2} \geq d^2, \tag{3.5}$$

and

$$A^\pm_{h,k}(d,\mathcal{X}) = \sum_{\substack{hm\equiv\pm kn \ (\mathrm{mod}\ d)\\ (mn,d)=1}} \frac{1}{m_1^{\frac12+\alpha} m_2^{\frac12+\beta} n_1^{\frac12+\gamma} n_2^{\frac12+\delta}} V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn}{\mathcal{X}}\right).$$

We partition the sum $A^+_{h,k}(d,\mathcal{X})$ into two cases depending on whether $hm < kn$ or $hm > kn$. When $hm < kn$, we apply Lemma 2.4 to rewrite it as

$$\sum_{\substack{hm\equiv kn_1n_2 \,(\mathrm{mod}\ d)\\ (mn_1n_2,d)=1}} \frac{\sigma_{\alpha,\beta}(m)}{m^{\frac{1}{2}}n_1^{\frac{1}{2}+\gamma}n_2^{\frac{1}{2}+\delta}}\left(\omega\left(\frac{n_1}{n_2}\right)+\omega\left(\frac{n_2}{n_1}\right)\right)V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn_1n_2}{\mathcal{X}}\right).$$

When $hm > kn$, we use instead the identity

$$\omega\left(\frac{m_1}{m_2}\right)+\omega\left(\frac{m_2}{m_1}\right)=1.$$

This decomposition splits $A^+_{h,k}(d,\mathcal{X})$ into four terms, obtained by permuting the shifts $h,k,\alpha,\beta,\cdots$, so that

$$A^+_{h,k}(d,\mathcal{X})=\mathcal{A}^+_{h,k}(\alpha,\beta,\gamma,\delta)+\mathcal{A}^+_{h,k}(\alpha,\beta,\delta,\gamma)+\mathcal{A}^+_{k,h}(\gamma,\delta,\alpha,\beta)+\mathcal{A}^+_{k,h}(\delta,\gamma,\beta,\alpha),$$

where

$$\mathcal{A}^+_{h,k}(\alpha,\beta,\gamma,\delta)=\sum_{\substack{hm\equiv kn_1n_2 \,(\mathrm{mod}\ d)\\ (mn_1n_2,d)=1}} \frac{\sigma_{\alpha,\beta}(m)}{m^{\frac{1}{2}}n_1^{\frac{1}{2}+\gamma}n_2^{\frac{1}{2}+\delta}}\omega\left(\frac{n_1}{n_2}\right)V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn_1n_2}{\mathcal{X}}\right).$$

We apply a similar decomposition to $A^-_{h,k}(d,\mathcal{X})$, but now using the identity

$$\omega\left(\frac{hm}{kn}\right)+\omega\left(\frac{kn}{hm}\right)=1$$

instead of the previous $hm < kn$ and $hm > kn$ splitting. This partitions $A^-_{h,k}(d,\mathcal{X})$ into four terms

$$A^-_{h,k}(d,\mathcal{X})=\mathcal{A}^-_{h,k}(\alpha,\beta,\gamma,\delta)+\mathcal{A}^-_{h,k}(\alpha,\beta,\delta,\gamma)+\mathcal{A}^-_{k,h}(\gamma,\delta,\alpha,\beta)+\mathcal{A}^-_{k,h}(\delta,\gamma,\beta,\alpha),$$

where

$$\mathcal{A}^-_{h,k}(\alpha,\beta,\gamma,\delta)=\sum_{\substack{hm\equiv -kn_1n_2 \,(\mathrm{mod}\ d)\\ (mn_1n_2,d)=1}} \frac{\sigma_{\alpha,\beta}(m)}{m^{\frac{1}{2}}n_1^{\frac{1}{2}+\gamma}n_2^{\frac{1}{2}+\delta}}\omega\left(\frac{hm}{kn_1n_2}\right)V_{\alpha,\beta,\gamma,\delta}\left(\frac{mn_1n_2}{\mathcal{X}}\right).$$

Applying the smooth dyadic partition to localize the variables m and n, we obtain

$$\mathcal{A}^\pm_{h,k}(\alpha,\beta,\gamma,\delta)=\sum_{M,N}B^\pm_{h,k}(\alpha,\beta,\gamma,\delta;M,N),$$

where the summands, $B^\pm_{h,k}$, inherit the structure of $\mathcal{A}^\pm_{h,k}$, but incorporate smooth test functions $W\left(\frac{m}{M}\right)$ and $W\left(\frac{n}{N}\right)$. The analysis of $B^\pm_{h,k}(M,N)$ bifurcates based on the relative sizes of hM and kN. The main terms arise from the balanced terms, in which hM and kN are close in magnitude. Conversely, when hM and kN differ significantly, these $B^\pm_{h,k}(M,N)$ (dubbed unbalanced terms) contribute only to the error terms.

For balanced terms, we demonstrate that

$$B^\pm_{h,k}(\alpha,\beta,\gamma,\delta;M,N)=(\text{Main term})_{M,N}+E^\pm_{h,k}(M,N),$$

where the main terms are too intricate to explicitly state here. Instead, we focus on the summation over all M, N. The following single upper bound (see Lemma 6.1)

$$(\text{Main term})_{M,N} \ll_j q^{j\varepsilon}(hk)^{-\frac{1}{2}} \left(\frac{hM}{kN}\right)^{\frac{1}{2}} \left(1 + \frac{kN}{hM}\right)^{-j}, \quad \text{for any } j > 0,$$

extends the summation to unbalanced terms with an acceptable error. These analyses are detailed in Section 6.

We estimate the error term $E^{\pm}_{h,k}(M,N)$ in Section 8, leveraging the upper bound for sums of Kloosterman sums established in Theorem 1.4. Let $\mathcal{E}^{\pm}_{h,k}(M,N)$ denote the total contribution of $E^{\pm}_{h,k}(M,N)$ to $\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta)$. The following estimate holds.

Theorem 3.1. *For $hM \ll kN$, we have*

$$\mathcal{E}^{\pm}_{h,k}(M,N) \ll q^{\varepsilon}\left(\frac{hk}{q}\right)^{\frac{1}{4}}\left(\frac{kN}{hM}\right)^{\frac{1}{4}} + q^{\varepsilon}\left(\frac{hk}{q}\right)^{\frac{1}{2}}\left(\frac{kN}{hM}\right)^{\frac{1}{2}}. \tag{3.6}$$

Let $\mathcal{E}^{\pm}(H,K,M,N)$ denote the total contribution of $E^{\pm}_{h,k}(M,N)$ to $\mathcal{M}(\alpha,\beta,\gamma,\delta)$, where h and k range over $[H,2H]\times[K,2K]$. Explicitly,

$$\mathcal{E}^{\pm}(H,K,M,N) = \sum_{H\le h\le 2H}\sum_{K\le k\le 2K}\frac{\alpha_h\overline{\alpha_k}}{h^{\frac{1}{2}}k^{\frac{1}{2}}}\mathcal{E}^{\pm}_{h,k}(M,N).$$

A trivial estimation of the summation using (3.6) yields the following bound.

Theorem 3.2. *For $HM \ll KN$, we have*

$$\mathcal{E}^{\pm}(H,K,M,N) \ll \frac{(HK)^{\frac{3}{4}}}{q^{\frac{1}{4}-\varepsilon}}\left(\frac{KN}{HM}\right)^{\frac{1}{4}} + \frac{HK}{q^{\frac{1}{2}-\varepsilon}}\left(\frac{KN}{HM}\right)^{\frac{1}{2}}. \tag{3.7}$$

Let $\mathscr{B}^{\pm}_{h,k}(M,N)$ denote the total contribution of $B^{\pm}_{h,k}(M,N)$ to $\mathcal{M}_{h,k}(\alpha,\beta,\gamma,\delta)$, and let $\mathscr{B}^{\pm}(H,K,M,N)$ denote the total contribution of $B^{\pm}_{h,k}(M,N)$ to $\mathcal{M}(\alpha,\beta,\gamma,\delta)$, where h and k range over $[H,2H]\times[K,2K]$. When M is significantly smaller compared to N, we can bound them via the following trivial bounds

$$\mathscr{B}^{\pm}_{h,k}(M,N) \ll q^{-1+\varepsilon}(MN)^{\frac{1}{2}} + q^{\varepsilon}(M/N)^{\frac{1}{2}}, \quad \mathscr{B}^{\pm}(H,K,M,N) \ll q^{-1+\varepsilon}(HKMN)^{\frac{1}{2}}. \tag{3.8}$$

The proofs of Theorems 1.1–1.2 is completed by analyzing unbalanced terms in Sections 9–10, where $\mathscr{B}^{\pm}_{h,k}(M,N)$ and $\mathscr{B}^{\pm}(H,K,M,N)$ are reformulated as bilinear forms of incomplete Kloosterman sums. These are then analyzed using new bounds from [KSWX23] alongside Weil's well-known bound.

4. Some arithmetic sums

Lemma 4.1. *For complex numbers $\alpha,\beta,\gamma,\delta$ and any integer a, let $Y_a(\alpha,\beta,\gamma,\delta)$ be defined as in (1.7). We have*

$$Y_a(\alpha,\beta,\gamma,\delta) = Y_a(\beta,\alpha,\gamma,\delta) = Y_a(\beta,\alpha,\delta,\gamma) = Y_a(\alpha,\beta,\delta,\gamma).$$

Proof. Applying the identity

$$1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}} =\frac{1}{1-p^{\gamma-\delta}}\left(1-\frac{1}{p^{1+\alpha+\delta}}\right)\left(1-\frac{1}{p^{1+\beta+\delta}}\right) - \frac{p^{\gamma-\delta}}{1-p^{\gamma-\delta}}\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma}}\right),$$

we arrive at the expression

$$\begin{aligned} (4.1)\quad & Y_{p^k}(\alpha,\beta,\gamma,\delta)\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right) \\ &=\frac{1}{p^{k\gamma}}\left(\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right)+\frac{p^{\gamma-\delta}-p^{(k+1)(\gamma-\delta)}}{1-p^{\gamma-\delta}}\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma}}\right)\right) \\ &=\frac{p^{-(k+1)\gamma}}{p^{-\gamma}-p^{-\delta}}\left(1-\frac{1}{p^{1+\alpha+\delta}}\right)\left(1-\frac{1}{p^{1+\beta+\delta}}\right)+\frac{p^{-(k+1)\delta}}{p^{-\delta}-p^{-\gamma}}\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma}}\right). \end{aligned}$$

This expression is symmetric in α and β, as well as in γ and δ. Thus,

$$Y_{p^k}(\alpha,\beta,\gamma,\delta)=Y_{p^k}(\beta,\alpha,\gamma,\delta)=Y_{p^k}(\beta,\alpha,\delta,\gamma)=Y_{p^k}(\alpha,\beta,\delta,\gamma).$$

The general case follows from the multiplicativity of Y_a. □

The calculation of the diagonal terms requires evaluating the following arithmetic sum

$$\mathcal{F}_a(\alpha,\beta,\gamma,\delta)=\sum_{d|a^\infty}\frac{\sigma_{\alpha,\beta}(d)\sigma_{\gamma,\delta}(ad)}{d}.$$

Our next result evaluates $\mathcal{F}_a(\alpha,\beta,\gamma,\delta)$.

Lemma 4.2. *For complex numbers $\alpha,\beta,\gamma,\delta$ and any integer a, we have*

$$\begin{aligned}\mathcal{F}_a(\alpha,\beta,\gamma,\delta) =&Y_a(\alpha,\beta,\gamma,\delta)\prod_{p|a}\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\gamma}}\right)^{-1} \\ &\times\left(1-\frac{1}{p^{1+\alpha+\delta}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\delta}}\right)^{-1}\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right).\end{aligned}$$

Proof. By multiplicativity, it suffices to prove the case $a=p^k$, $k\geq 1$. Expanding $\sigma_{\alpha,\beta}(d)$ implies that

$$\mathcal{F}_{p^k}(\alpha,\beta,\gamma,\delta)=\sum_{i=0}^{\infty}\frac{1}{p^{i(1+\alpha)}}\sum_{j=0}^{\infty}\frac{1}{p^{j(1+\beta)}}\sigma_{\gamma,\delta}(p^{k+i+j}).$$

Applying the identity

$$\sigma_{\gamma,\delta}(p^{k+i+j})=\frac{p^{-(k+i+j+1)\gamma}-p^{-(k+i+j+1)\delta}}{p^{-\gamma}-p^{-\delta}},$$

we see that

$$\begin{aligned}\mathcal{F}_{p^k}(\alpha,\beta,\gamma,\delta) =&\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\gamma}}\right)^{-1}\frac{p^{-(k+1)\gamma}}{p^{-\gamma}-p^{-\delta}}\\ &-\left(1-\frac{1}{p^{1+\alpha+\delta}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\delta}}\right)^{-1}\frac{p^{-(k+1)\delta}}{p^{-\gamma}-p^{-\delta}}.\end{aligned}$$

By (4.1), this simplifies to

$$\begin{aligned}\mathcal{F}_{p^k}(\alpha,\beta,\gamma,\delta) =&Y_{p^k}(\alpha,\beta,\gamma,\delta)\left(1-\frac{1}{p^{1+\alpha+\delta}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\delta}}\right)^{-1}\\ &\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)^{-1}\left(1-\frac{1}{p^{1+\beta+\gamma}}\right)^{-1}\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right).\end{aligned}$$

The general case follows from multiplicativity. □

Our next lemma deals with an arithmetic sum of $Y_a(\alpha,\beta,\gamma,\delta)$. For notational convenience, we define

(4.2) $$\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s) = Y_h(-\delta-s,\beta+s,\gamma+s,-\alpha-s)Y_k(\gamma+s,-\alpha-s,-\delta-s,\beta+s).$$

Lemma 4.3. *Let q be a square-free integer and*

$$\mathcal{G}_q(\alpha,\beta,\gamma,\delta,s) = \sum_{\substack{abc|q^2\\(a,b)=1}}\frac{\varpi_{\alpha,\beta}(ac,q)\varpi_{\gamma,\delta}(bc,q)}{a^{\frac12+s}b^{\frac12+s}c^{2s}}\mathcal{Y}_{a,b,\alpha,\beta,\gamma,\delta}(s).$$

Then we have

$$\begin{aligned}\mathcal{G}_q(\alpha,\beta,\gamma,\delta,s) =& \prod_{p|q}\left(1-\frac{1}{p^{1-\alpha+\beta}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma+2s}}\right)\\ &\times\left(1-\frac{1}{p^{1+\gamma-\delta}}\right)\left(1-\frac{1}{p^{2-\alpha+\beta+\gamma-\delta}}\right)^{-1}\left(1-\frac{2}{p}+\frac{1}{p^{1+\gamma+\delta+2s}}\right).\end{aligned}$$

Proof. By multiplicativity, it suffices to prove the case $q = p$ (prime). From (3.5) and (4.2), we have

$$\begin{aligned}\varpi_{\alpha,\beta}(p,p) &= -\frac{1}{p^{\frac12+\alpha}}-\frac{1}{p^{\frac12+\beta}}, \quad \varpi_{\alpha,\beta}(p^2,p) = \frac{1}{p^{1+\alpha+\beta}},\\ \mathcal{Y}_{p^k,1,\alpha,\beta,\gamma,\delta}(s) &= Y_{p^k}(-\delta-s,\beta+s,\gamma+s,-\alpha-s),\\ \mathcal{Y}_{1,p^k,\alpha,\beta,\gamma,\delta}(s) &= Y_{p^k}(\gamma+s,-\alpha-s,-\delta-s,\beta+s).\end{aligned}$$

Additionally,

$$Y_p(\alpha,\beta,\gamma,\delta)\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right) = \frac{1}{p^\gamma}+\frac{1}{p^\delta}-\frac{1}{p^{1+\alpha+\gamma+\delta}}-\frac{1}{p^{1+\beta+\gamma+\delta}},$$

and

$$Y_{p^2}(\alpha,\beta,\gamma,\delta)\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right)=\frac{1}{p^{2\gamma}}\left(1-\frac{1}{p^{2+\alpha+\beta+\gamma+\delta}}\right)$$
$$+\left(\frac{1}{p^{\gamma+\delta}}+\frac{1}{p^{2\delta}}\right)\left(1-\frac{1}{p^{1+\alpha+\gamma}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma}}\right).$$

Expanding $\mathcal{G}_p$ renders that

$$\begin{aligned}\mathcal{G}_p(\alpha,\beta,\gamma,\delta,s)&=1+\frac{\varpi_{\alpha,\beta}(p,p)}{p^{\frac12+s}}\mathcal{Y}_{p,1,\alpha,\beta,\gamma,\delta}(s)+\frac{\varpi_{\gamma,\delta}(p,p)}{p^{\frac12+s}}\mathcal{Y}_{1,p,\alpha,\beta,\gamma,\delta}(s)\\&+\frac{\varpi_{\alpha,\beta}(p^2,p)}{p^{1+2s}}\mathcal{Y}_{p^2,1,\alpha,\beta,\gamma,\delta}(s)+\frac{\varpi_{\gamma,\delta}(p^2,p)}{p^{1+2s}}\mathcal{Y}_{1,p^2,\alpha,\beta,\gamma,\delta}(s)\\&+\frac{\varpi_{\alpha,\beta}(p^2,p)\varpi_{\gamma,\delta}(p,p)}{p^{\frac12+3s}}\mathcal{Y}_{p,1,\alpha,\beta,\gamma,\delta}(s)+\frac{\varpi_{\alpha,\beta}(p,p)\varpi_{\gamma,\delta}(p^2,p)}{p^{\frac12+3s}}\mathcal{Y}_{1,p,\alpha,\beta,\gamma,\delta}(s)\\&+\frac{\varpi_{\alpha,\beta}(p,p)\varpi_{\gamma,\delta}(p,p)}{p^{2s}}+\frac{\varpi_{\alpha,\beta}(p^2,p)\varpi_{\gamma,\delta}(p^2,p)}{p^{4s}}.\end{aligned}$$

A straightforward calculation yields

$$\mathcal{G}_p(\alpha,\beta,\gamma,\delta,s)\left(1-\frac{1}{p^{2-\alpha+\beta+\gamma-\delta}}\right)=\left(1-\frac{1}{p^{1-\alpha+\beta}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma+2s}}\right)$$
$$\times\left(1-\frac{1}{p^{1+\gamma-\delta}}\right)\left(1-\frac{2}{p}+\frac{1}{p^{1+\gamma+\delta+2s}}\right),$$

which can be verified using a mathematical software, such as *Mathematica*. The general case follows from multiplicativity. □

Our last two lemmas evaluate certain Dirichlet series.

Lemma 4.4. *Let a,b be coprime integers. For $Re(s)>1$ and $Re(\lambda)>-1$, we have*

$$\sum_{r\ge1}\frac{1}{r^s}\sum_{(l,b)=1}\frac{c_{al}(r)}{l^{2+\lambda}}=a^{1-s}\prod_{p|a}\left(1-\frac{1}{p^{1-s}}\right)\frac{\zeta(s)\zeta_b(1+\lambda+s)}{\zeta_{ab}(2+\lambda)}.$$

Proof. Let $r\to a_1b_1r'$ and $l\to a_2l'$ with $a_1=(a^\infty,r)$, $a_2=(a^\infty,l)$, and $b_1=(b^\infty,r)$. This factorization ensures

$$c_{al}(r)=c_{aa_2}(a_1)c_{l'}(r').$$

Thus, the sum decomposes as

$$\sum_{r\ge1}\frac{1}{r^s}\sum_{(l,b)=1}\frac{c_{al}(r)}{l^{2+\lambda}}=C(a)\prod_{p|b}\left(1-\frac{1}{p^s}\right)^{-1}\sum_{(r,ab)=1}\frac{1}{r^s}\sum_{(l,ab)=1}\frac{c_l(r)}{l^{2+\lambda}},\tag{4.3}$$

where

$$C(a)=\sum_{a_1|a^\infty}\sum_{a_2|a^\infty}\frac{c_{aa_2}(a_1)}{a_1^sa_2^{2+\lambda}}.$$

By [Wu23+a, Lemma 6.2], the inner sums equal to

$$\sum_{(r,ab)=1} \frac{1}{r^s} \sum_{(l,ab)=1} \frac{c_l(r)}{l^{2+\lambda}} = \frac{\zeta_{ab}(s)\zeta_{ab}(1+\lambda+s)}{\zeta_{ab}(2+\lambda)}.$$

To compute $C(a)$, it suffices to consider $a = p^k$ by multiplicativity. Using the identities for Ramanujan sums

$$c_{p^j}(p^i) = \begin{cases} 0, & \text{for } i < j-1, \\ -p^{j-1} & \text{for } i = j-1, \\ \varphi(p^j) & \text{for } i \geq j. \end{cases}$$

we write $a_1 = p^i$, $a_2 = p^j$ and partition the sum for $C(p^k)$ into two terms, based on whether $i = k+j-1$ or $i \geq k+j$. Explicitly,

$$\begin{aligned} C(p^k) &= \sum_{j=0}^{\infty} \frac{-p^{k+j-1}}{p^{j(2+\lambda)+(k+j-1)s}} + \sum_{j=0}^{\infty} \frac{\varphi(p^{k+j})}{p^{j(2+\lambda)}} \sum_{i=k+j}^{\infty} \frac{1}{p^{is}} \\ &= \frac{-p^s}{p^{ks-k+1}} \left(1 - \frac{1}{p^{1+\lambda+s}}\right)^{-1} + \frac{p-1}{p^{ks-k+1}} \left(1 - \frac{1}{p^s}\right)^{-1} \left(1 - \frac{1}{p^{1+\lambda+s}}\right)^{-1} \\ &= \frac{1-p^{s-1}}{p^{k(s-1)}} \left(1 - \frac{1}{p^s}\right)^{-1} \left(1 - \frac{1}{p^{1+\lambda+s}}\right)^{-1}. \end{aligned}$$

It then follows from multiplicativity that for any a,

$$C(a) = a^{1-s} \prod_{p|a} \left(1 - \frac{1}{p^{1-s}}\right)\left(1 - \frac{1}{p^s}\right)^{-1} \left(1 - \frac{1}{p^{1+\lambda+s}}\right)^{-1}.$$

Substituting this back into (4.3) yields the lemma. $\square$

Lemma 4.5. *For any $z \in \mathbb{C}$, we have*

$$\sum_{d|q} \frac{\varphi(d)^2}{d^{1+z}} \mu\left(\frac{q}{d}\right) \prod_{p|q_d} \left(1 - \frac{2}{p} + \frac{1}{p^{1+z}}\right) = \frac{\varphi^*(q)}{q^z} \prod_{p|q} \left(1 - \frac{1}{p^{1-z}}\right).$$

Proof. By multiplicativity, it suffices to verify the identity for $q = p$ (prime) and $q = p^k$ $(k > 1)$.

For $q = p$, the left-hand side (LHS) simplifies to

$$\mu(p)\left(1 - \frac{2}{p} + \frac{1}{p^{1+z}}\right) + \frac{\varphi(p)^2}{p^{1+z}} = \frac{p-2}{p^z}\left(1 - \frac{1}{p^{1-z}}\right),$$

which is the same as the right-hand side (RHS) of the desired identity in this case.

For $q = p^k$ with $k > 1$, the LHS reduces to

$$\sum_{d|p^k} \frac{\varphi(d)^2}{d^{1+z}} \mu\left(\frac{p^k}{d}\right).$$

Note that here the only contributions come from $d = p^{k-1}$ and $d = p^k$, so that

$$\mu(p)\frac{\varphi(p^{k-1})^2}{p^{(k-1)(1+z)}} + \frac{\varphi(p^k)^2}{p^{k(1+z)}} = \frac{\varphi(p^k)^2}{p^{k(1+z)}}\left(1 - \frac{1}{p^{1-z}}\right).$$

One checks directly that right-hand side matches this expression. This therefore completes the proof. □

5. Diagonal terms

Note that when considering the contribution from the diagonal terms $hm = kn$, the summation over d is simplified by the factor $\varphi^*(q)$. Let $h_1 = (h^\infty, m)$ and $k_1 = (k^\infty, n)$. Given that $(h,k) = 1$, we have

$$\begin{aligned}\mathscr{M}_{h,k}^{D} =&(hk)^{-\frac{1}{2}} \sum_{\substack{h_1|h^\infty\\ k_1|k^\infty}} \frac{\sigma_{\alpha,\beta}(h_1kk_1)\sigma_{\gamma,\delta}(hh_1k_1)}{h_1k_1} \sum_{(n,hkq)=1} \frac{\sigma_{\alpha,\beta}(n)\sigma_{\gamma,\delta}(n)}{n} V\left(\frac{hkh_1^2k_1^2n^2}{q^2}\right)\\ =&\frac{1}{2\pi i}\int_{(1)} \frac{G(s)}{s}\frac{q^{2s}g(s)}{(hk)^{\frac{1}{2}+s}} \sum_{\substack{h_1|h^\infty\\ k_1|k^\infty}} \frac{\sigma_{\alpha,\beta}(h_1kk_1)\sigma_{\gamma,\delta}(hh_1k_1)}{h_1^{1+2s}k_1^{1+2s}} \sum_{(n,hkq)=1} \frac{\sigma_{\alpha,\beta}(n)\sigma_{\gamma,\delta}(n)}{n^{1+2s}} \mathrm{d}s.\end{aligned}$$

By the Ramanujan identity, the summation over n evaluates to

$$\frac{\zeta_{hkq}(1+\alpha+\gamma+2s)\zeta_{hkq}(1+\alpha+\delta+2s)\zeta_{hkq}(1+\beta+\gamma+2s)\zeta_{hkq}(1+\beta+\delta+2s)}{\zeta_{hkq}(2+\alpha+\beta+\gamma+\delta+4s)}.$$

We shift the contour of integration to $\mathrm{Re}(s) = -\frac{1}{4}+\varepsilon$, encountering only a simple pole at $s = 0$ since all potential poles from the zeta functions are canceled by zeros of $G(s)$. The remaining integral contributes an error term of $O\left(q^{-\frac{1}{2}+\varepsilon}\right)$.

For the residue at $s = 0$, the sums involving h_1 and k_1 correspond to $\mathcal{F}_h(\alpha,\beta,\gamma,\delta)$ and $\mathcal{F}_k(\gamma,\delta,\alpha,\beta)$ respectively. Applying Lemma 4.2 and the definition of $Z_{h,k,q}$ given in (1.6), we conclude that

$$\begin{aligned}\mathscr{M}_{h,k}^{D}(\alpha,\beta,\gamma,\delta) =&(hk)^{-\frac{1}{2}} Y_h(\alpha,\beta,\gamma,\delta)Y_k(\gamma,\delta,\alpha,\beta)Z_q(\alpha,\beta,\gamma,\delta) + O\left(q^{-\frac{1}{2}+\varepsilon}\right)\\ =&(hk)^{-\frac{1}{2}} Z_{h,k,q}(\alpha,\beta,\gamma,\delta) + O\left(q^{-\frac{1}{2}+\varepsilon}\right).\end{aligned} \tag{5.1}$$

We evaluate $\mathscr{M}_{\overline{h},k}^{D}(\alpha,\beta,\gamma,\delta)$ via (3.2) to see that

$$\mathscr{M}_{\overline{h},k}^{D}(\alpha,\beta,\gamma,\delta) = (hk)^{-\frac{1}{2}} X_{\alpha,\beta,\gamma,\delta} Z_{h,k,q}(-\gamma,-\delta,-\alpha,-\beta) + O\left(q^{-\frac{1}{2}+\varepsilon}\right). \tag{5.2}$$

6. Evaluation of $B_{h,k}^{\pm}$

Note that the condition $(n_2, d) = 1$ in $B_{h,k}^{\pm}$ is automatically satisfied by other variables. We factorize $h = h_1h_2$ where $h_2 = (h, n_1)$. By canceling h_2 in the congruence condition and performing the variable substitution $n_1 \to h_2n_1$, we transform the congruence into

$$kn_1n_2 = dr \pm h_1m \tag{6.1}$$

for integers $r \geq 1$. We define the following functions

$$F^{+}(m) = \frac{1}{(dr+h_1m)^{\frac{1}{2}+\delta}}\omega\left(\frac{h_2kn_1^2}{dr+h_1m}\right)V\left(\frac{h_2m(dr+h_1m)}{k\mathcal{X}}\right), \tag{6.2}$$

and

$$F^{-}(m) = \frac{1}{(dr-h_1m)^{\frac{1}{2}+\delta}}\omega\left(\frac{h_2kn_1^2}{dr-h_1m}\right)\omega\left(\frac{h_1m}{dr-h_1m}\right)V\left(\frac{h_2m(dr-h_1m)}{k\mathcal{X}}\right). \tag{6.3}$$

We now eliminate n_2 using (6.1) to see that

$$B^{\pm}_{h,k} = \sum_{h_1h_2=h} \frac{k^{\frac12+\delta}}{h_2^{\frac12+\gamma}} \sum_{r\geq 1} \sum_{(n_1,dh_1)=1} n_1^{\delta-\gamma} \sum_{\substack{h_1m\equiv\mp dr \,(\mathrm{mod}\ kn_1)\\(m,d)=1}} \frac{\sigma_{\alpha-\beta}(m)}{m^{\frac12+\alpha}} F^{\pm}_{M,N}(m),$$

where

$$F^{\pm}_{M,N}(m) = F^{\pm}(m) W\left(\frac{m}{M}\right) W\left(\frac{h_2(dr\pm h_1m)}{kN}\right). \tag{6.4}$$

For any divisor $g \mid kn_1$, we factorize $g = k_1 l$ where $l = (g, n_1)$ and $k_1 \mid k$. This establishes a bijection between g and the ordered pairs (k_1, l). Leveraging this decomposition, we eliminate the congruence condition using primitive additive characters

$$\begin{aligned}\mathbf{1}_{h_1m\equiv\mp dr\,(\mathrm{mod}\ kn_1)} &= \frac{1}{kn_1} \sum_{g\mid kn_1} \sideset{}{^*}\sum_{t\,(\mathrm{mod}\ g)} e\left(\frac{drt\pm h_1mt}{g}\right)\\ &= \frac{1}{kn_1} \sum_{k_1\mid k} \sum_{\substack{fl=n_1\\(f,k_1)=1}} \sideset{}{^*}\sum_{t\,(\mathrm{mod}\ k_1l)} e\left(\frac{drt\pm h_1mt}{k_1l}\right).\end{aligned}$$

Substituting this into the expression for $B^{\pm}_{h,k}$, we obtain that

$$\begin{aligned}B^{\pm}_{h,k} = \frac{1}{k^{\frac12-\delta}} \sum_{h_1h_2=h} \sum_{k_1\mid k} \frac{1}{h_2^{\frac12+\gamma}}\\ \times \sum_{r\geq1} \sum_{(n_1,dh_1)=1} \frac{1}{n_1^{1+\gamma-\delta}} \sum_{\substack{fl=n_1\\(f,k_1)=1}} \sideset{}{^*}\sum_{t\,(\mathrm{mod}\ k_1l)} e\left(\frac{drt}{k_1l}\right) D_d\left(\tfrac12+\alpha, \alpha-\beta, \frac{\pm h_1 t}{k_1 l}; F^{\pm}_{M,N}\right).\end{aligned} \tag{6.5}$$

By Lemma 2.6, the dominant terms of D_d are given by

$$\frac{\varphi(d)}{d}\frac{\zeta_d(1-\alpha+\beta)}{(k_1l)^{1-\alpha+\beta}} \widetilde{F}^{\pm}_{M,N}\left(\tfrac12-\alpha\right), \quad \frac{\varphi(d)}{d}\frac{\zeta_d(1+\alpha-\beta)}{(k_1l)^{1+\alpha-\beta}} \widetilde{F}^{\pm}_{M,N}\left(\tfrac12-\beta\right).$$

Substituting these into (6.5), we denote by $P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N)$ the contribution of the first term above, while noting that the sum over t there evaluates to $c_{k_1l}(r)$. By symmetry, we obtain

$$B^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;h,k) = P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N) + P^{\pm}_{h,k}(\beta,\alpha,\gamma,\delta;M,N) + E^{\pm}_{h,k}(M,N),$$

where

$$\begin{aligned}P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N) = &\frac{\varphi(d)}{d}\frac{\zeta_q(1-\alpha+\beta)}{k^{\frac12-\delta}} \sum_{h_1h_2=h} \sum_{k_1\mid k} \frac{1}{h_2^{\frac12+\gamma}} \frac{1}{k_1^{1-\alpha+\beta}}\\ &\times \sum_{r\geq1} \sum_{(f,dh_1k_1)=1} \frac{1}{f^{1+\gamma-\delta}} \sum_{(l,dh_1)=1} \frac{c_{k_1l}(r)}{l^{2-\alpha+\beta+\gamma-\delta}} \widetilde{F}^{\pm}_{M,N}\left(\tfrac12-\alpha\right).\end{aligned} \tag{6.6}$$

We focus on analyzing $P^{\pm}_{h,k}$ in the remainder of this section and defer the treatment of the error term $E^{\pm}_{h,k}$ to Section 8.

6.1. **Upper bound for main terms.** For the terms $P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N)$, we establish the following upper bound that permits extension of the main terms to all M and N.

Lemma 6.1. *For any $j>0$, we have*

$$P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N) \ll_j q^{j\varepsilon}\frac{(MN)^{\frac{1}{2}}}{d}\left(1+\frac{kN}{hM}\right)^{-j}.$$

Proof. Observe from (6.4) that for any $j>0$,

$$F^{\pm}_{M,N}(m) \ll_j q^{j\varepsilon}\left(\frac{h_2}{kN}\right)^{\frac{1}{2}}\left(1+\frac{kN}{hM}\right)^{-j}, \tag{6.7}$$

where the dominant factor follows directly from (6.2) and (6.3). For $F^{+}_{M,N}(m)$, the rapid decay in (6.7) originates from the W-function in (6.4) with $r\geq 1$, while for $F^{-}_{M,N}(m)$ it arises from the second ω-function in (6.3). This rapid decay property also implies that the number of terms r is $O(kN/h_2 d)$.

We denote the Mellin transform $\widetilde{F}^{\pm}_{M,N}$ of $F^{\pm}_{M,N}$ by

$$\widetilde{F}^{\pm}_{M,N}(u) = \int_0^\infty F^{\pm}_{M,N}(x)x^{u-1}\mathrm{d}x.$$

Applying (6.7) to the above yields

$$\widetilde{F}^{\pm}_{M,N}\left(\tfrac{1}{2}-\alpha\right), \widetilde{F}^{\pm}_{M,N}\left(\tfrac{1}{2}-\beta\right) \ll_j q^{j\varepsilon}\left(\frac{h_2M}{kN}\right)^{\frac{1}{2}}\left(1+\frac{kN}{hM}\right)^{-j}.$$

Substituting this into (6.6) and performing trivial summation with the number of r not exceeding $O(kN/h_2d)$ allow us to complete the proof. □

We now apply Theorem 3.1 and Lemma 6.1 to extend the summation range of the main terms to include unbalanced terms with controlled error terms. After summing over all M and N using (2.1), we proceed with the final analyses of the main terms. We define

$$\mathcal{P}^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta) = \sum_{M,N} P^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta;M,N).$$

It follows from (3.4) that its contribution to $\mathscr{M}^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta)$ is given by

$$\begin{aligned} Q^{\pm}_{h,k}(\alpha,\beta,\gamma,\delta) = \frac{1}{\varphi^*(q)}\sum_{d|q}\varphi(d)\mu\left(\frac{q}{d}\right) \\ \times \sum_{\substack{abc|q_d^2\\(a,b)=1}} \varpi_{\alpha,\beta}(ac,q_d)\varpi_{\gamma,\delta}(bc,q_d)\mathcal{P}^{\pm}_{ah,bk}(\alpha,\beta,\gamma,\delta), \end{aligned} \tag{6.8}$$

which will be evaluated in the following sections.

6.2. **Main terms with negative sign.** We compute $\mathcal{P}^-_{h,k}$ and subsequently $\mathcal{Q}^-_{h,k}$ in this section. The following lemma furnishes an explicit expression for $\mathcal{P}^-_{h,k}$.

Lemma 6.2. *With $c_s = \frac{1}{4}$, $c_v = \varepsilon$, we have*

$$\mathcal{P}^-_{h,k}(\alpha,\beta,\gamma,\delta) = \frac{\varphi(d)}{d}(hk)^{-\frac{1}{2}}\zeta_d(1-\alpha+\beta)\zeta_d(1+\gamma-\delta)\frac{1}{(2\pi i)^2}\int_{(c_s)}\int_{(c_v)}\frac{X^s}{d^{\alpha+\delta+2s}}$$
$$\times H^-(s,v)\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\frac{\zeta(\alpha+\delta+2s)\zeta_d(1+\beta+\gamma+2s)}{\zeta_d(2-\alpha+\beta+\gamma-\delta)}\mathrm{d}s\mathrm{d}v + O\left(\frac{k^{\frac{1}{4}}X^{\frac{1}{2}+\varepsilon}}{dh^{\frac{1}{4}}}\right),$$

where $\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)$ is defined as in (4.2), and

$$H^-(s,v) = \frac{G(s)}{s}g(s)\widetilde{\omega}(v)\frac{\Gamma(\frac{1}{2}-\alpha-s-v)\Gamma(\frac{1}{2}-\delta-s+v)}{\Gamma(1-\alpha-\delta-2s)}.$$

Proof. We apply the partition identity (2.1) to see that

$$\mathcal{P}^-_{h,k}(\alpha,\beta,\gamma,\delta) = \frac{\varphi(d)}{d}\frac{\zeta_d(1-\alpha+\beta)}{k^{\frac{1}{2}-\delta}}\sum_{h_1h_2=h}\sum_{k_1|k}\frac{1}{h_2^{\frac{1}{2}+\gamma}}\frac{1}{k_1^{1-\alpha+\beta}}$$
$$\times\sum_{r\geq 1}\sum_{(f,dh_1k_1)=1}\frac{1}{f^{1+\gamma-\delta}}\sum_{(l,dh_1)=1}\frac{c_{k_1l}(r)}{l^{2-\alpha+\beta+\gamma-\delta}}\widetilde{F}^-\left(\tfrac{1}{2}-\alpha\right). \tag{6.9}$$

We analytically separate the variables r, f and l in F^- by taking Mellin transforms of ω and V, obtaining

$$F^-(x) = \frac{1}{(2\pi i)^3}\int_{(c_s)}\int_{(c_u)}\int_{(c_v)}\frac{X^sk^{s-u}(dr-h_1x)^{-\frac{1}{2}-\delta-s+u+v}}{h_2^{u+s}h_1^vf^{2u}l^{2u}x^{s+v}}\widetilde{\omega}(u)\widetilde{\omega}(v)\frac{G(s)}{s}g(s)\mathrm{d}s\mathrm{d}u\mathrm{d}v.$$

Initially setting $c_s = c_u = c_v = \varepsilon$, we further separate variables in $(dr-h_1x)^{-\frac{1}{2}-\delta-s+u+v}$ using the following formula

$$(1-x)^{-b} = \frac{1}{2\pi i}\int_{(c_w)}\frac{\Gamma(w)\Gamma(1-b)}{\Gamma(1-b+w)}x^{-w}\mathrm{d}w, \tag{6.10}$$

which is valid for $0 < x, c_w, \mathrm{Re}(b) < 1$.

This then leads to

$$F^-(x) = \frac{1}{(2\pi i)^4}\int_{(c_s)}\int_{(c_u)}\int_{(c_v)}\int_{(c_w)}\frac{X^sk^{s-u}(dr)^{-\frac{1}{2}-\delta-s+u+v+w}}{h_2^{s+u}h_1^{v+w}f^{2u}l^{2u}x^{s+v+w}}$$
$$\times\frac{G(s)}{s}g(s)\widetilde{\omega}(u)\widetilde{\omega}(v)\frac{\Gamma(w)\Gamma(\frac{1}{2}-\delta-s+u+v)}{\Gamma(\frac{1}{2}-\delta-s+u+v+w)}\mathrm{d}s\mathrm{d}u\mathrm{d}v\mathrm{d}w,$$

which allows us to write its Mellin transform as

$$\widetilde{F}^-\left(\tfrac{1}{2}-\alpha\right) = \frac{1}{(2\pi i)^3}\int_{(c_s)}\int_{(c_u)}\int_{(c_v)}\frac{X^sk^{s-u}(dr)^{-\alpha-\delta-2s+u}}{h_2^{s+u}h_1^{\frac{1}{2}-\alpha-s}f^{2u}l^{2u}}H_1^-(s,u,v)\mathrm{d}s\mathrm{d}u\mathrm{d}v, \tag{6.11}$$

where

$$H_1^-(s,u,v) = \frac{G(s)}{s}g(s)\widetilde{\omega}(u)\widetilde{\omega}(v)\frac{\Gamma(\frac{1}{2}-\alpha-s-v)\Gamma(\frac{1}{2}-\delta-s+u+v)}{\Gamma(1-\alpha-\delta-2s+u)}.$$

Substituting (6.11) into (6.9) yields

$$\mathcal{P}^{-}_{h,k} = \frac{\varphi(d)}{d}\frac{\zeta_q(1-\alpha+\beta)}{h^{\frac12}k^{\frac12}}$$
$$\times \frac{1}{(2\pi i)^3}\int_{(c_s)}\int_{(c_u)}\int_{(c_v)} H_1^-(s,u,v)\frac{X^s k^{\delta+s-u}}{d^{\alpha+\delta+2s-u}h^{\gamma+s+u}}\sum_{h_1|h}\sum_{k_1|k}\frac{h_1^{\alpha+\gamma+2s+u}}{k_1^{1-\alpha+\beta}}$$
$$\times \sum_{(f,dh_1k_1)=1}\frac{1}{f^{1+\gamma-\delta+2u}}\sum_{r\geq 1}\frac{1}{r^{\alpha+\delta+2s-u}}\sum_{(l,dh_1)=1}\frac{c_{k_1 l}(r)}{l^{2-\alpha+\beta+\gamma-\delta+2u}}\mathrm{d}s\mathrm{d}u\mathrm{d}v.$$

We shift the contours of the integral to $c_s = \frac12+\varepsilon$ and $c_u = \varepsilon$, crossing the only simple pole at $s = \frac12 - \alpha - v$. We first evaluate the new contour integral, leaving the residue for later.

For the new contour integral, the sums over f, l, and r converge absolutely. The f-sum evaluates to

$$\prod_{p|h_1k_1}\left(1-\frac{1}{p^{1+\gamma-\delta+2u}}\right)\zeta_d(1+\gamma-\delta+2u).$$

By Lemma 4.4, the innermost two sums over l and r simplify to

$$k_1^{1-\alpha-\delta-2s+u}\prod_{p|k_1}\left(1-\frac{1}{p^{1-\alpha-\delta-2s+u}}\right)\frac{\zeta(\alpha+\delta+2s-u)\zeta_{dh_1}(1+\beta+\gamma+2s+u)}{\zeta_{dh_1k_1}(2-\alpha+\beta+\gamma-\delta+2u)}.$$

After these simplifications, the factor involving h in the integral becomes

$$\begin{aligned}(6.12)\quad &\frac{1}{h^{\gamma+s+u}}\sum_{h_1|h}h_1^{\alpha+\gamma+2s+u}\prod_{p|h_1}\left(1-\frac{1}{p^{1+\gamma-\delta+2u}}\right)\left(1-\frac{1}{p^{1+\beta+\gamma+2s+u}}\right)\left(1-\frac{1}{p^{2-\alpha+\beta+\gamma-\delta+2u}}\right)^{-1}\\ &= Y_h(-\delta-s+u,\beta+s,\gamma+s+u,-\alpha-s).\end{aligned}$$

Similarly, the factor involving k simplifies to

$$\begin{aligned}(6.13)\quad &k^{\delta+s-u}\sum_{k_1|k}\frac{1}{k_1^{\beta+\delta+2s-u}}\prod_{p|k_1}\left(1-\frac{1}{p^{1+\gamma-\delta+2u}}\right)\left(1-\frac{1}{p^{1-\alpha-\delta-2s+u}}\right)\left(1-\frac{1}{p^{2-\alpha+\beta+\gamma-\delta+2u}}\right)^{-1}\\ &= Y_k(\gamma+s+u,-\alpha-s,-\delta-s+u,\beta+s).\end{aligned}$$

In conclusion, the new contour integral evaluates to

$$\begin{aligned}(6.14)\quad &\frac{\varphi(d)}{d}(hk)^{-\frac12}\zeta_d(1-\alpha+\beta)\frac{1}{(2\pi i)^3}\int_{(c_s)}\int_{(c_u)}\int_{(c_v)}\frac{X^s}{d^{\alpha+\delta+2s-u}}H_1^-(s,u,v)\\ &\times Y_h(-\delta-s+u,\beta+s,\gamma+s+u,-\alpha-s)Y_k(\gamma+s+u,-\alpha-s,-\delta-s+u,\beta+s)\\ &\times\frac{\zeta_d(1+\gamma-\delta+2u)\zeta(\alpha+\delta+2s-u)\zeta_d(1+\beta+\gamma+2s+u)}{\zeta_d(2-\alpha+\beta+\gamma-\delta+2u)}\mathrm{d}s\mathrm{d}u\mathrm{d}v.\end{aligned}$$

We then shift the contour of the integral to $c_s = \frac14$, crossing the pole at $s = \frac12 - \alpha - v$ again. Note here that the pole of $\zeta(\alpha+\delta+2s-u)$ is canceled out by the zero of $H_1^-(s,u,v)$.

Now that we have crossed the pole at $s = \frac12 - \alpha - v$ twice, we observe that the two residues ultimately cancel each other out. Let us examine this cancellation in detail.

Clearly, the expression simplifies to

$$\operatorname{Res}_{s=\frac{1}{2}-\alpha-v} H_1^-(s,u,v) = \frac{G\left(\frac{1}{2}-\alpha-v\right)}{\frac{1}{2}-\alpha-v} g\left(\frac{1}{2}-\alpha-v\right)\widetilde{\omega}(u)\widetilde{\omega}(v),$$

As $G\left(\frac{1}{2}-\alpha\right) = 0$, it follows that for any fixed $u \neq 0$ the function $\operatorname{Res}_{s=\frac{1}{2}-\alpha-v} H_1^-(s,u,v)$ remains analytic in the half-plane $\operatorname{Re}(v) < \frac{1}{2} - \operatorname{Re}(\alpha)$. For the first residue computation, we can take c_v to be negative, thereby guaranteeing absolute convergence for all sums over f, l, and r. Upon evaluating these sums as previously done, it becomes evident that the two residues at $s = \frac{1}{2} - \alpha - v$ annihilate each other.

We proceed to shift the contours of the remaining integral to $c_s = \frac{1}{4}$ and $c_v = \frac{1}{4} - \varepsilon$, followed by $c_u = -\frac{1}{2} + 2\varepsilon$. During this process, we encounter the only pole at $u = 0$, keeping in mind that the pole of $\zeta(1+\gamma-\delta+2u)$ is canceled out by $\widetilde{\omega}\left(\frac{\delta-\gamma}{2}\right) = 0$. The main contribution comes from the residue at $u = 0$. For the integral on the new lines, analysis of (6.12) and (6.13) implies that

$$Y_h(-\delta-s+u, \beta+s, \gamma+s+u, -\alpha-s) \ll h^{\frac{1}{4}+\varepsilon}, \;\; Y_k(\gamma+s+u, -\alpha-s, -\delta-s+u, \beta+s) \ll k^{\frac{3}{4}+\varepsilon}.$$

A straightforward estimation then shows that the integral on the new lines is bounded by $O\left(\frac{k^{\frac{1}{4}}X^{\frac{1}{2}+\varepsilon}}{dh^{\frac{1}{4}}}\right)$. Therefore, we obtain the final expression

$$\begin{aligned}\mathcal{P}_{h,k}^- =& \frac{\varphi(d)}{d}(hk)^{-\frac{1}{2}}\zeta_d(1-\alpha+\beta)\zeta_d(1+\gamma-\delta)\frac{1}{(2\pi i)^2}\int_{(c_s)}\int_{(c_v)}\frac{X^s}{d^{\alpha+\delta+2s}}H^-(s,v)\\ &\times \mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\frac{\zeta(\alpha+\delta+2s)\zeta_d(1+\beta+\gamma+2s)}{\zeta_d(2-\alpha+\beta+\gamma-\delta)}\mathrm{d}s\mathrm{d}v + O\left(\frac{k^{\frac{1}{4}}X^{\frac{1}{2}+\varepsilon}}{dh^{\frac{1}{4}}}\right).\end{aligned}$$

This completes the proof of the lemma. □

We now apply Lemma 6.2 to (6.8) to derive the following expression for $Q_{h,k}^-$.

Proposition 6.3. *With $c_s = \frac{1}{4}$, $c_v = \varepsilon$, we have*

$$\begin{aligned}Q_{h,k}^-(\alpha,\beta,\gamma,\delta) =& (hk)^{-\frac{1}{2}}\frac{\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\frac{1}{2\pi i}\int_{(c_s)}\frac{G(s)}{s}g(s)\mathcal{H}_{\alpha,\delta}^-(s)\\ &\times \mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\zeta_q(1-\alpha-\delta-2s)\zeta_q(1+\beta+\gamma+2s)\mathrm{d}s + O\left(h^{-\frac{1}{4}}k^{\frac{1}{4}}q^{-\frac{1}{2}+\varepsilon}\right),\end{aligned}$$

where

$$\begin{aligned}\mathcal{H}_{\alpha,\delta}^-(s) =& \pi^{2s-\frac{1}{2}}\left(\frac{q}{\pi}\right)^{-\alpha-\delta}\frac{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}\\ &\times \frac{1}{2\pi i}\int_{(c_v)}\frac{\Gamma\left(\frac{1}{2}-\alpha-s-v\right)\Gamma\left(\frac{1}{2}-\delta-s+v\right)}{\Gamma(1-\alpha-\delta-2s)}\widetilde{\omega}(v)dv.\end{aligned}$$

Proof. We begin by substituting the expression of $\mathcal{P}_{ah,bk}^-(\alpha,\beta,\gamma,\delta)$, obtained from Lemma 6.2, into the formula (6.8). The resulting sum over a, b, c transforms into $\mathcal{G}_{q_d}(\alpha,\beta,\gamma,\delta,s)$

through the multiplicativity of $\mathcal{Y}$. Given that q_d is square-free due to the presence of $\mu(q/d)$, we apply Lemma 4.3 to $\mathcal{G}_{q_d}$ to see that

$$\begin{aligned}\mathcal{Q}^-_{h,k} =&(hk)^{-\frac{1}{2}}\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)\frac{1}{(2\pi i)^2}\int_{(c_s)}\int_{(c_v)} q^{2s}H^-(s,v)\\ &\times\frac{1}{\varphi^*(q)}\sum_{d|q}\frac{\varphi(d)^2}{d^{1+\alpha+\delta+2s}}\mu\left(\frac{q}{d}\right)\prod_{p|q_d}\left(1-\frac{2}{p}+\frac{1}{p^{1+\alpha+\delta+2s}}\right)\\ &\times\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\frac{\zeta(\alpha+\delta+2s)\zeta_q(1+\beta+\gamma+2s)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\mathrm{d}s\mathrm{d}v+O\left(h^{-\frac{1}{4}}k^{\frac{1}{4}}q^{-\frac{1}{2}+\varepsilon}\right).\end{aligned}$$

Next, we evaluate the sum over d using Lemma 4.5, yielding

$$\begin{aligned}\mathcal{Q}^-_{h,k} =&(hk)^{-\frac{1}{2}}\frac{\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\frac{q^{-\alpha-\delta}}{(2\pi i)^2}\int_{(c_s)}\int_{(c_v)}H^-(s,v)\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\\ &\times\prod_{p|q}\left(1-\frac{1}{p^{1-\alpha-\delta-2s}}\right)\zeta(\alpha+\delta+2s)\zeta_q(1+\beta+\gamma+2s)\mathrm{d}s\mathrm{d}v+O\left(h^{-\frac{1}{4}}k^{\frac{1}{4}}q^{-\frac{1}{2}+\varepsilon}\right).\end{aligned}$$

Finally, we apply the functional equation

$$\prod_{p|q}\left(1-\frac{1}{p^{1-\alpha-\delta-2s}}\right)\zeta(\alpha+\delta+2s)=\frac{1}{\pi^{\frac{1}{2}-\alpha-\delta-2s}}\frac{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}\zeta_q(1-\alpha-\delta-2s)$$

to establish the proposition. □

6.3. **Main terms with positive sign.** The proofs for $\mathcal{P}^+_{h,k}$ and $\mathcal{Q}^+_{h,k}$ follow from analogous arguments to the negative sign case, subject to appropriate modifications. This next lemma constitutes a positive-sign counterpart to Lemma 6.2.

Lemma 6.4. *With $c_s=\frac{1}{2}+\varepsilon$, we have*

$$\begin{aligned}\mathcal{P}^+_{h,k}(\alpha,\beta,\gamma,\delta)=&\frac{\varphi(d)}{d}(hk)^{-\frac{1}{2}}\zeta_d(1-\alpha+\beta)\zeta_d(1+\gamma-\delta)\frac{1}{2\pi i}\int_{(c_s)}\frac{\mathcal{X}^s}{d^{\alpha+\delta+2s}}H^+(s)\\ &\times\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\frac{\zeta(\alpha+\delta+2s)\zeta_d(1+\beta+\gamma+2s)}{\zeta_d(2-\alpha+\beta+\gamma-\delta)}\mathrm{d}s+O\left(\frac{k^{\frac{1}{4}}\mathcal{X}^{\frac{1}{2}+\varepsilon}}{dh^{\frac{1}{4}}}\right),\end{aligned}$$

where $\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)$ is defined as in (4.2), and

$$H^+(s)=\frac{G(s)}{s}g(s)\frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\alpha+\delta+2s)}{\Gamma(\frac{1}{2}+\delta+s)}.$$

Proof. The analysis closely follows that of the proof of Lemma 6.2, with several modifications made below. We use the identity

$$(1+x)^{-b}=\frac{1}{2\pi i}\int_{(c_w)}\frac{\Gamma(w)\Gamma(b-w)}{\Gamma(b)}x^{-w}\mathrm{d}w$$

for $0 < c_w < \mathrm{Re}(b)$ in place of (6.10). Since there is only one ω-function in F^+, the gamma factor evaluates to

$$H_1^+(s,u) = \frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\alpha+\delta+2s-u)}{\Gamma(\frac{1}{2}+\delta+s-u)}\frac{G(s)}{s}g(s)\widetilde{\omega}(u).$$

Here the pole of $\Gamma(\frac{1}{2}-\alpha-s)$ is canceled out by $G(\frac{1}{2}-\alpha)=0$, and we shift the s-integral to $c_s=\frac{1}{2}+\varepsilon$ to ensure absolute convergence, without crossing any pole. After performing all summations, we obtain an analogy of (6.14). The u-contour is then shifted to $c_u=-\frac{1}{2}+\varepsilon$, crossing only the pole at $u=0$. The residue yields the main term of $\mathcal{P}^+_{h,k}$, while the remaining integrals are bound by the error term using identical arguments as the negative sign case. □

We now apply Lemma 6.4 to (6.8) to obtain the corresponding expression for $Q^+_{h,k}$.

Proposition 6.5. *With $c_s=\frac{1}{4}$, we have*

$$\begin{aligned}Q^+_{h,k}(\alpha,\beta,\gamma,\delta) = (hk)^{-\frac{1}{2}}\frac{\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\frac{1}{2\pi i}\int_{(c_s)}\frac{G(s)}{s}g(s)H^+_{\alpha,\delta}(s)\\ \times\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\zeta_q(1-\alpha-\delta-2s)\zeta_q(1+\beta+\gamma+2s)\mathrm{d}s + O\left((hk)^{-\frac{1}{2}}q^{-\frac{1}{2}+\varepsilon}\right),\end{aligned}$$

where

$$H^+_{\alpha,\delta}(s) = \pi^{2s-\frac{1}{2}}\left(\frac{q}{\pi}\right)^{-\alpha-\delta}\frac{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}\frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\alpha+\delta+2s)}{\Gamma(\frac{1}{2}+\delta+s)}.$$

Proof. This proof follows along the same line of argument as Proposition 6.3, with Lemma 6.4 substituted for Lemma 6.2. The final step involves shifting the contour of integration from $c_s=\frac{1}{2}+\varepsilon$ to $c_s=\frac{1}{4}$, which introduces no additional residues since the potential poles at $s=\frac{1}{2}(1-\alpha-\delta)$ and $s=\frac{1}{2}-\alpha$ of $H^+_{\alpha,\delta}(s)$ are precisely canceled by zeros of $\zeta_q(1-\alpha-\delta-2s)$ and $G(s)$ respectively. □

7. Assembling the main terms

We establish the following combination

$$\begin{aligned}(7.1)\qquad & Q^+_{h,k}(\alpha,\beta,\gamma,\delta)+Q^+_{k,h}(\delta,\gamma,\beta,\alpha)+Q^-_{h,k}(\alpha,\beta,\gamma,\delta)+Q^-_{k,h}(\delta,\gamma,\beta,\alpha)\\ &=Q_{h,k}(\alpha,\beta,\gamma,\delta)+O\left(h^{-\frac{1}{4}}k^{\frac{1}{4}}q^{-\frac{1}{2}+\varepsilon}\right).\end{aligned}$$

This yields the full contribution of main terms to $\mathscr{M}_{h,k}$ as

$$Q_{h,k}(\alpha,\beta,\gamma,\delta)+Q_{h,k}(\beta,\alpha,\gamma,\delta)+Q_{h,k}(\alpha,\beta,\delta,\gamma)+Q_{h,k}(\beta,\alpha,\delta,\gamma).$$

Lemma 7.1. *We have*

$$\begin{aligned}Q_{h,k}(\alpha,\beta,\gamma,\delta) =&(hk)^{-\frac{1}{2}}\frac{\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\frac{1}{2\pi i}\int_{(\frac{1}{4})}\frac{G(s)}{s}g(s)H_{\alpha,\delta}(s)\\ &\times\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s)\zeta_q(1-\alpha-\delta-2s)\zeta_q(1+\beta+\gamma+2s)\mathrm{d}s,\end{aligned}$$

where

$$H_{\alpha,\delta}(s) = \pi^{2s-\frac{1}{2}}\left(\frac{q}{\pi}\right)^{-\alpha-\delta}\frac{\Gamma\left(\frac{\frac{1}{2}-\alpha-s}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\alpha+s}{2}\right)}\frac{\Gamma\left(\frac{\frac{1}{2}-\delta-s}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\delta+s}{2}\right)}.$$

Proof. From (4.2) and Lemma 4.1, we immediately obtain the symmetry relation

$$\mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s) = \mathcal{Y}_{k,h,\delta,\gamma,\beta,\alpha}(s).$$

Consequently, by Proposition 6.3 and Proposition 6.5, all terms on the left-hand side of (7.1) share identical main contribution up to gamma function factors. Specifically, we have

$$H^-_{\alpha,\delta}(s) + H^-_{\delta,\alpha}(s) = \pi^{2s-\frac{1}{2}}\left(\frac{q}{\pi}\right)^{-\alpha-\delta}\frac{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}\frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\frac{1}{2}-\delta-s)}{\Gamma(1-\alpha-\delta-2s)}.$$

The complete gamma factor combination is then given by

$$\begin{aligned}H_{\alpha,\delta}(s) &= H^+_{\alpha,\delta}(s) + H^+_{\delta,\alpha}(s) + H^-_{\alpha,\delta}(s) + H^-_{\delta,\alpha}(s)\\ &= \pi^{2s-\frac{1}{2}}\left(\frac{q}{\pi}\right)^{-\alpha-\delta}\frac{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}\Gamma_{\alpha,\delta}(s),\end{aligned}$$

where

$$\begin{aligned}\Gamma_{\alpha,\delta}(s) =&\frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\alpha+\delta+2s)}{\Gamma(\frac{1}{2}+\delta+s)}\\ &+\frac{\Gamma(\frac{1}{2}-\alpha-s)\Gamma(\frac{1}{2}-\delta-s)}{\Gamma(1-\alpha-\delta-2s)} + \frac{\Gamma(\frac{1}{2}-\delta-s)\Gamma(\alpha+\delta+2s)}{\Gamma(\frac{1}{2}+\alpha+s)}.\end{aligned}$$

The proof is completed by verifying the identity

$$\Gamma_{\alpha,\delta}(s) = \pi^{\frac{1}{2}}\frac{\Gamma\left(\frac{\alpha+\delta+2s}{2}\right)}{\Gamma\left(\frac{1-\alpha-\delta-2s}{2}\right)}\frac{\Gamma\left(\frac{\frac{1}{2}-\alpha-s}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\alpha+s}{2}\right)}\frac{\Gamma\left(\frac{\frac{1}{2}-\delta-s}{2}\right)}{\Gamma\left(\frac{\frac{1}{2}+\delta+s}{2}\right)},$$

which follows from standard gamma function trigonometric identities as established in [You11, Lemma 8.2]. □

The main term of $\mathscr{M}_{\overline{h,k}}$ is then determined by the relation (3.2) as follows

$$Q_{\overline{h,k}}(\alpha,\beta,\gamma,\delta) + Q_{\overline{h,k}}(\beta,\alpha,\gamma,\delta) + Q_{\overline{h,k}}(\alpha,\beta,\delta,\gamma) + Q_{\overline{h,k}}(\beta,\alpha,\delta,\gamma),$$

where the individual components are defined by

$$Q_{\overline{h,k}}(\alpha,\beta,\gamma,\delta) = X_{\alpha,\beta,\gamma,\delta}Q_{k,h}(-\alpha,-\beta,-\gamma,-\delta). \tag{7.2}$$

Lemma 7.2. *We have*

$$Q_{h,k}(\alpha,\beta,\gamma,\delta) + Q_{\overline{h,k}}(\beta,\alpha,\delta,\gamma) = (hk)^{-\frac{1}{2}}Z_{h,k,q}(-\delta,\beta,\gamma,-\alpha).$$

Proof. By Lemma 7.1 and the relationship (7.2), we obtain the explicit expression

$$Q_{\overline{h,k}}(\beta,\alpha,\delta,\gamma) = \frac{X_{\alpha,\beta,\gamma,\delta}}{(hk)^{\frac{1}{2}}}\frac{\zeta_q(1-\alpha+\beta)\zeta_q(1+\gamma-\delta)}{\zeta_q(2-\alpha+\beta+\gamma-\delta)}\frac{1}{2\pi i}\int_{(-\frac{1}{4})}\frac{G(s)}{s}g_{-\alpha,-\beta,-\gamma,-\delta}(-s)$$
$$\times H_{-\beta,-\gamma}(-s)\mathcal{Y}_{k,h,-\beta,-\alpha,-\delta,-\gamma}(-s)\zeta_q(1-\alpha-\delta+2s)\zeta_q(1+\beta+\gamma-2s)\mathrm{d}s.$$

From (4.2), we deduce the symmetry property

$$\mathcal{Y}_{k,h,-\beta,-\alpha,-\delta,-\gamma}(-s) = \mathcal{Y}_{h,k,\alpha,\beta,\gamma,\delta}(s).$$

Furthermore, it is known that

$$X_{\alpha,\beta,\gamma,\delta}g_{-\alpha,-\beta,-\gamma,-\delta}(-s)H_{-\beta,-\gamma}(-s) = g_{\alpha,\beta,\gamma,\delta}(s)H_{\alpha,\delta}(s),$$

which corresponds precisely to the last formula in [You11, p.39].

Applying these two relations yields that $Q_{h,k}(\alpha,\beta,\gamma,\delta)$ and $Q_{\overline{h,k}}(\beta,\alpha,\delta,\gamma)$ share identical integrands but are integrated along conjugate contours $c_s = \frac{1}{4}$ and $c_s = -\frac{1}{4}$ respectively. The lemma then follows immediately from Cauchy's theorem. □

The arguments in Lemma 7.2 can be similary applied to other pairs of $Q_{h,k}$ and $Q_{\overline{h,k}}$. Thus that the total contribution of the sum of all of them equals

$$\begin{aligned}&X_{\alpha,\gamma}Z_{h,k,q}(\beta,-\gamma,\delta,-\alpha) + X_{\beta,\gamma}Z_{h,k,q}(\alpha,-\gamma,\delta,-\beta) + X_{\alpha,\delta}Z_{h,k,q}(\beta,-\delta,\gamma,-\alpha)\\ &\quad + X_{\beta,\delta}Z_{h,k,q}(\alpha,-\delta,\gamma,-\beta) + O\left(h^{-\frac{1}{4}}k^{\frac{1}{4}}q^{-\frac{1}{2}+\varepsilon}\right).\end{aligned} \tag{7.3}$$

8. Error terms for balanced terms

For the purpose of simplifying subsequent formulae, we uniformly set all shift parameters $\alpha,\beta,\gamma,\delta$ to 0. All arguments can be straightforwardly extended to handle sufficiently small nonzero values $\alpha,\beta,\gamma,\delta$ without introducing additional complexity.

We want to apply Lemma 2.6 to the auxiliary function

$$\mathcal{V}(m,r,l) = \left(\frac{kN}{h_2}\right)^{\frac{1}{2}} F^{\pm}_{M,N}(m),$$

where we set $n_1 = fl$ in the definitions (6.2), (6.3) and (6.4). This function inherits the type $W(m/M)$ characteristic from $F^{\pm}_{M,N}(m)$, while simultaneously being smooth in variables r and l. It satisfies the uniform bound

$$m^{j_1}r^{j_2}l^{j_3}\frac{\partial^{j_1+j_2+j_3}}{\partial m^{j_1}\partial r^{j_2}\partial l^{j_3}}\mathcal{V}(m,r,l) \ll_{j_1,j_2,j_3} q^{(j_1+j_2+j_3)\varepsilon}.$$

Applying the above for the functions $\mathring{\mathcal{V}}_{\pm}$ defined in Lemma 2.6, we obtain the corresponding estimates

$$x^{j_1}r^{j_2}l^{j_3}\frac{\partial^{j_1+j_2+j_3}}{\partial x^{j_1}\partial r^{j_2}\partial l^{j_3}}\mathring{\mathcal{V}}_{\pm}(x,r,l) \ll_{j_1,j_2,j_3} q^{(j_1+j_2+j_3)\varepsilon}M^{\frac{1}{2}}.$$

Since the treatments for $E^{+}_{h,k}(M,N)$ and $E^{-}_{h,k}(M,N)$ are completely parallel, we unify their notation by writing $E_{h,k}(M,N)$ while suppressing the sign distinction. The condition

$(l, h_1 d) = 1$ in (6.5) decomposes into two independent conditions $(l, h_1 a) = 1$ and $(l, d_a) = 1$. We eliminate the latter using the Möbius formula

$$\mathbf{1}_{(l,d_a)=1} = \sum_{c|(l,d_a)} \mu(c).$$

After applying dyadic decomposition to localize the variable n_1 in (6.5) ($n_1 \asymp N_1$ with $h_2 N_1 \ll N^{\frac{1}{2}}$), we apply Lemma 2.6 to (6.5) to obtain

$$E_{h,k}(M,N) \ll q^{\varepsilon}(hk)^{-\frac{1}{2}}\left(\frac{hM}{kN}\right)^{\frac{1}{2}} \max_{(f,h_1,k_1,a,c)}(|E_+| + |E_-|), \tag{8.1}$$

where

$$E_{\pm} = \frac{1}{a^2c^2k_1}\sum_r \alpha_r \sum_m \beta_m \sum_{(l,h_1a)=1} \frac{1}{l^2}\Omega(r,m,l) S(\overline{h_1a^2}dr, \pm m; ck_1 l),$$

with $\alpha_r \ll r^{\varepsilon}$, $\beta_m \ll m^{\varepsilon}$ and the parameters subject to the constraints

$$f \le N_1, \quad h_1h_2 = h, \quad k_1 \mid k, \quad (hk,d) = 1, \quad ac \mid d, \quad (a,c) = 1.$$

Here, we also applied dyadic decomposition to localize all variables r, m, and l, and $\Omega(r,m,l)$ is a product of $\mathring{\mathcal{V}}_{\pm}(m,r,l)$ with the three W-functions: $W(m/M^*)$, $W(r/R)$, and $W(l/L)$.

We set

$$Q = uv, \quad u = ck_1, \quad v = h_1a^2, \quad s = d,$$

which implies the simplified bound

$$M^* \ll \frac{(Lu)^2 v}{h_1 M}.$$

The sums $E_{\pm}$ can be reformulated as

$$E_{\pm} = \frac{h_1k_1}{u^2v}\sum_r \alpha_r \sum_m \beta_m \sum_{(l,v)=1} \frac{1}{l^2}\Omega(r,m,l) S(sr\overline{v}, \pm m; lu),$$

where $\Omega(r,m,l)$ inherits smoothness properties from $\mathring{\mathcal{V}}_{\pm}(m,r,l)$ and W-functions, having compact support in $[R,2R]\times[M^*,2M^*]\times[L,2L]$ with

$$R \ll \frac{kN}{dh_2}, \quad M^* \ll \frac{(Lack_1)^2}{M}, \quad L = \frac{N_1}{cf}, \quad N_1 \ll \frac{N^{\frac{1}{2}}}{h_2}.$$

Applying Theorem 1.4, we obtain the estimation

$$E_{\pm} \ll \frac{h_1k_1}{s^{\frac{1}{2}}}\frac{(1+X^{-\vartheta})X}{1+X}\left(1+X+\frac{s^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1+X+\frac{(s,Q)^{\frac{1}{2}}R}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1+X+\frac{(M^*)^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right),$$

where the auxiliary parameter X satisfies

$$X = \frac{(sRM^*)^{\frac{1}{2}}}{uv^{\frac{1}{2}}L} \ll \left(\frac{kN}{hM}\right)^{\frac{1}{2}}. \tag{8.2}$$

Note that we have the following relations.

$$\frac{M^*}{Q} \ll \frac{L^2u}{h_1M} \ll \frac{N_1^2k_1}{h_1M} \ll \frac{kN}{hM}, \quad \frac{(s,Q)}{Q} \ll \frac{1}{h_1k_1}.$$

Substituting the estimation for X given in (8.2) and simplifying, we derive that

$$E_{\pm} \ll \frac{h_1k_1}{s^{\frac{1}{2}}}\left(\frac{kN}{hM}+\frac{s}{h_1k_1}\right)^{\frac{1}{4}}\left(\frac{kN}{hM}+\frac{k^2N^2}{dh_1k_1^2}\right)^{\frac{1}{4}}\left(\frac{kN}{hM}\right)^{\frac{1}{2}}.$$

Applying this to (8.1), we obtain the final bound

$$E_{h,k}(M,N) \ll q^{\varepsilon}\left(\frac{hk}{d}\right)^{\frac{1}{2}}\left(\frac{(kN)^{\frac{1}{2}}}{(hM)^{\frac{1}{2}}}+\frac{k^{\frac{1}{2}}N^{\frac{3}{4}}}{h^{\frac{1}{2}}M^{\frac{1}{4}}d^{\frac{1}{2}}}+\frac{(dN)^{\frac{1}{4}}}{h^{\frac{1}{2}}M^{\frac{1}{4}}}+\frac{N^{\frac{1}{2}}}{(hd)^{\frac{1}{2}}}\right).$$

This result, when substituted into (3.4), completes the proof of (3.6).

9. Unbalanced terms

In this section, we preliminary treat the unbalanced terms $B^{\pm}_{h,k}(M,N)$, preparing for the subsequent proof of theorems in next section. For simplicity, we again set all parameters $\alpha,\beta,\gamma,\delta$ to 0. The arguments can be carried out straightforwardly to handle sufficiently small nonzero values of $\alpha,\beta,\gamma,\delta$ without additional cost.

Without loss of generality, we adopt the convention $hM \ll kN$ throughout this section. Given the identical treatment of $B^{+}_{h,k}(M,N)$ and $B^{-}_{h,k}(M,N)$, we focus exclusively on the former, denoting its contribution to $\mathcal{M}_{h,k}$ by $\mathscr{B}_{h,k}(M,N)$. The corresponding contribution to $\mathcal{M}(\alpha,\beta,\gamma,\delta)$ is given by

$$\mathscr{B}(H,K,M,N)=\frac{1}{\sqrt{HK}}\sum_{\substack{h\leq H\\(hk,q)=1}}\sum_{k\leq K}\alpha_h\beta_k\mathscr{B}_{h,k}(M,N), \tag{9.1}$$

where $\alpha_h \ll h^{\varepsilon}$, $\beta_k \ll k^{\varepsilon}$ are some complex coefficients supported on dyadic intervals $[H,2H]$ and $[K,2K]$, respectively.

We employ the Poisson summation formula to transform the unbalanced terms $\mathscr{B}_{h,k}(M,N)$ and $\mathscr{B}(H,K,M,N)$ into bilinear forms of incomplete Kloosterman sums. To derive sharp estimates for them, we supplement Weil's bound with the following specialized estimate.

Lemma 9.1 ([KSWX23, Theorem 2.4])**.** *Let q be a positive integer and α_a,β_b be sequences of complex numbers. For any $A,B\geq 1$, the estimate*

$$\sum_{a\leq A}\alpha_a\left|\sum_{\substack{b\leq B\\(b,q)=1}}\beta_b e\left(\frac{ca\overline{b}}{q}\right)\right| \ll |\boldsymbol{\alpha}|_2|\boldsymbol{\beta}|_{\infty}A^{\frac{1}{2}}Bq^{\varepsilon}\left(A^{-\frac{1}{2}}B^{-\frac{1}{4}}q^{\frac{1}{4}}+A^{-\frac{1}{2}}+q^{-\frac{1}{2}}+B^{-\frac{1}{2}}\right)$$

holds uniformly in c with $(c,q)=1$.

This lemma provides a refined control over the bilinear interactions, which is essential for handling the unbalanced regime.

We now eliminate the functions ω and V via Mellin transforms, then apply dyadic decomposition to localize the variables n_1,n_2 within the ranges $n_1 \asymp N_1$ and $n_2 \asymp N_2$, where $N_1N_2 \asymp N$ with $N_1 \leq N_2$. This procedure effectively reduces the term $\mathscr{B}_{h,k}(M,N)$

to the following structured form

$$(9.2)\qquad \mathscr{B}_{h,k}(M,N) = \frac{1}{\varphi^*(q)}\sum_{d|q}\varphi(d)\mu\left(\frac{q}{d}\right)\frac{1}{\sqrt{MN}} \times \sum_{(m,q)=1}\sum_{(n_1,q)=1}\tau(m)W\left(\frac{m}{M}\right)W\left(\frac{n_1}{N_1}\right)\sum_{\substack{n_2\equiv\overline{kn_1}hm\,(\mathrm{mod}\ d)\\ (n_2,q_d)=1}} W\left(\frac{n_2}{N_2}\right)$$

for some smooth function W. After removing the condition $(n_2, q_d) = 1$ using the Möbius function, we apply the Poisson summation formula to the sum over n_2, obtaining

$$\begin{aligned}\sum_{\substack{n_2\equiv\overline{kn_1}hm\,(\mathrm{mod}\ d)\\ (n_2,q_d)=1}} W\left(\frac{n_2}{N_2}\right) &= \sum_{g|q_d}\mu(g)\sum_{\substack{n_2\equiv\overline{kn_1}hm\,(\mathrm{mod}\ d)\\ n_2\equiv 0\,(\mathrm{mod}\ g)}} W\left(\frac{n_2}{N_2}\right)\\ &= \frac{N_2}{d}\sum_{g|q_d}\frac{\mu(g)}{g}\sum_{l} e\left(\frac{hlm\overline{kgn_1}}{d}\right)\widehat{W}\left(\frac{lN_2}{gd}\right).\end{aligned}$$

The "main term" arises from the $l = 0$ term, which yields

$$\widehat{W}(0)\frac{N_2}{d}\sum_{g|q_d}\frac{\mu(g)}{g}.$$

Substituting this into (9.2), we observe that the contribution vanishes due to the following identity

$$\sum_{d|q}\frac{\varphi(d)}{d}\mu\left(\frac{q}{d}\right)\sum_{g|q_d}\frac{\mu(g)}{g} = 0.$$

This identity is readily verified for prime powers $q = p^k$ ($k \geq 1$) and extends to general q by multiplicativity. Consequently, the remaining contribution is captured by the $l \neq 0$ terms, leading to

$$(9.3)\qquad \mathscr{B}_{h,k}(M,N) \ll \frac{q^{\varepsilon}}{\varphi^*(q)}\sum_{d|q}\sum_{g|q_d}\frac{\varphi(d)}{dg}\frac{N_2}{\sqrt{MN}} \times\left|\sum_{l\neq 0}\sum_{(m,q)=1}\sum_{(n_1,q)=1}\tau(m)e\left(\frac{hlm\overline{kgn_1}}{d}\right)W\left(\frac{m}{M}\right)W\left(\frac{n_1}{N_1}\right)\widehat{W}\left(\frac{lN_2}{gd}\right)\right|.$$

10. Proof of Theorems 1.1–1.2

10.1. Proof of Theorem 1.1. We adopt the exponential parametrization

$$h = q^{\varsigma},\quad k = q^{\varrho},\quad M = q^{\mu},\quad N = q^{\nu},$$

where the unbalance condition ensures $\mu + \varsigma$ and $\nu + \varrho$ differ significantly.

We deduce from (3.1), (3.3), (3.6), (5.1), (5.2) and (7.3), that the desired relation in (1.8) holds, provided that we show for $\eta \in (0, \frac{1}{20}(1 - 6\varrho))$,

$$(10.1)\qquad \mathscr{B}_{h,k}(M,N) \ll q^{-\eta+\varepsilon}.$$

Without loss of generality, we may restrict to $(l, d) = 1$ in (9.3). Applying Weil's bound to the sum over n_1 and estimating the remaining terms trivially yields

$$\mathscr{B}_{h,k}(M, N) \ll q^{\varepsilon} \sum_{d|q} \frac{d}{\varphi^*(q)} \left(\frac{M}{N}\right)^{\frac{1}{2}} \left(d^{\frac{1}{2}} + \frac{N_1}{d}\right) \ll q^{\frac{1}{2}+\varepsilon} \left(\frac{M}{N}\right)^{\frac{1}{2}}, \tag{10.2}$$

where the final inequality follows from $N_1 \ll N^{\frac{1}{2}} \ll q^{1+\varepsilon}$.

Applying Lemma 9.1 with parameters $a = lm$ and $b = n_1$ to (9.3), we obtain that

$$\mathscr{B}_{h,k}(M, N) \ll q^{\varepsilon} \sum_{d|q} \frac{d}{\varphi^*(q)} \left(\frac{MN_1}{N_2}\right)^{\frac{1}{2}} \left(\left(\frac{Md}{N_2}\right)^{-\frac{1}{2}} N_1^{-\frac{1}{4}} d^{\frac{1}{4}} + \left(\frac{Md}{N_2}\right)^{-\frac{1}{2}} + d^{-\frac{1}{2}} + N_1^{-\frac{1}{2}}\right).$$

After simplification, this reduces to

$$\mathscr{B}_{h,k}(M, N) \ll q^{\varepsilon}\left(\left(\frac{N_1}{q}\right)^{\frac{1}{4}} + \left(\frac{M}{N}\right)^{\frac{1}{2}} N_1^{\frac{1}{2}}\right). \tag{10.3}$$

Define $N_1 = q^{\nu_1}$. We deduce from (3.6), (3.8) and (10.2) that it suffices to establish (10.1) for the parameters satisfying

$$2 - 2\eta \le \mu + \nu \le 2, \quad 1 - 2\varrho - 4\eta \le \nu - \mu \le 1 + 2\eta, \quad \nu_1 \le \tfrac{1}{2}\nu. \tag{10.4}$$

From (10.4), we immediately deduce that

$$\nu_1 \le \tfrac{1}{4}((\mu + \nu) + (\nu - \mu)) \le \tfrac{3}{4} + \tfrac{1}{2}\eta. \tag{10.5}$$

Applying (10.3) yields

$$\mathscr{B}_{h,k}(M, N) \ll q^{-\frac{1}{4}(1-\nu_1)+\varepsilon} + q^{-\frac{1}{2}(\nu-\mu)+\frac{1}{2}\nu_1},$$

while by (10.4) and (10.5), we see that

$$\begin{aligned} -\tfrac{1}{4}(1 - \nu_1) &\le -\tfrac{1}{4}(\tfrac{1}{4} - \tfrac{1}{2}\eta) \le -\eta \quad \text{for} \quad \eta \le \tfrac{1}{18}, \\ -\tfrac{1}{2}(\nu - \mu) + \tfrac{1}{2}\nu_1 &\le -\tfrac{1}{2}(\nu - \mu) + \tfrac{1}{8}\left((\mu + \nu) + (\nu - \mu)\right) \\ &\le -\tfrac{1}{8} + \tfrac{3}{4}\varrho + \tfrac{3}{2}\eta \le -\eta \quad \text{for} \quad \eta \le \tfrac{1}{20}(1 - 6\varrho). \end{aligned}$$

The above readily implies that validity of (10.1) subject to the conditions given in (10.4). This completes the proof of Theorem 1.1.

10.2. **Proof of Theorem 1.2.** We apply (9.1) and (10.2) to obtain an initial estimate for $\mathscr{B}(H, K, M, N)$ so that

$$\mathscr{B}(H, K, M, N) \ll q^{\varepsilon}\left(\frac{qHKM}{N}\right)^{\frac{1}{2}}. \tag{10.6}$$

To proceed further, we apply (9.3) to see that

$$\mathscr{B}(H, K, M, N) \ll \frac{q^{\varepsilon}}{\varphi^*(q)} \sum_{d|q} \sum_{g|q_d} \frac{\varphi(d)}{gd} \frac{N_2}{\sqrt{HKMN}}$$
$$\times \left| \sum_{h\le H} \sum_{\substack{k\le K \\ (ab,q)=1}} \alpha_h \beta_k \sum_{l\neq 0} \sum_{(m,q)=1} \sum_{n_1} \tau(m) e\left(\frac{hlm\overline{kgn_1}}{d}\right) W\left(\frac{m}{M}\right) W\left(\frac{n_1}{N_1}\right) \widehat{W}\left(\frac{lN_2}{gd}\right) \right|.$$

Applying Lemma 9.1 with $a = hlm$ and $b = kn_1$, we obtain

$$\mathscr{B}(H,K,M,N) \ll q^{\varepsilon}\left(\left(\frac{KN_1}{q}\right)^{\frac{1}{4}} + \left(\frac{KN_1}{q}\right)^{\frac{1}{2}}\right) + \left(\frac{qHM}{KN}\right)^{\frac{1}{2}}\left(\left(\frac{KN_1}{q}\right)^{\frac{1}{2}} + \frac{KN_1}{q}\right). \tag{10.7}$$

We now define the exponential parameters

$$H = q^{\varsigma}, \quad K = q^{\varrho}, \quad M = q^{\mu}, \quad N = q^{\nu}, \quad N_1 = q^{\nu_1}.$$

We infer from (3.7), (3.8) and (10.6) that it is enough to establish the bound

$$\mathscr{B}(H,K,M,N) \ll q^{-\eta+\varepsilon} \tag{10.8}$$

for $\eta \in (0, \frac{1}{20}(1-12\varrho-10\varsigma))$, subject to the constraints

$$2-\varrho-\varsigma-2\eta \le \mu+\nu \le 2, \quad 1-4\varrho-2\varsigma-4\eta \le \nu-\mu \le 1+\varrho+\varsigma+2\eta, \quad \nu_1 \le \tfrac{1}{2}\nu. \tag{10.9}$$

We first derive from (10.9) that

$$\nu_1 \le \tfrac{1}{4}((\mu+\nu)+(\nu-\mu)) \le \tfrac{3}{4} + \tfrac{1}{4}\varrho + \tfrac{1}{4}\varsigma + \tfrac{1}{2}\eta, \tag{10.10}$$

which implies

$$\frac{KN_1}{q} = q^{-(1-\nu_1-\varrho)} \ll 1,$$

provided that $0 < \eta < \frac{1}{2}(1-5\varrho-\varsigma)$. Substituting this into (10.7) yields

$$\mathscr{B}(H,K,M,N) \ll q^{-\frac{1}{4}(1-\nu_1-\varrho)+\varepsilon} + q^{-\frac{1}{2}(\nu-\mu)+\frac{1}{2}\nu_1+\frac{1}{2}\varsigma}.$$

We then apply (10.9) and (10.10) to see that

$$\begin{aligned}
&-\tfrac{1}{4}(1-\nu_1-\varrho) \le -\tfrac{1}{4}(\tfrac{1}{4}-\tfrac{5}{4}\varrho-\tfrac{1}{4}\varsigma-\tfrac{1}{2}\eta) \le -\eta \quad \text{for} \quad \eta \le \tfrac{1}{18}(1-5\varrho-\varsigma),\\
&-\tfrac{1}{2}(\nu-\mu)+\tfrac{1}{2}\nu_1+\tfrac{1}{2}\varsigma \le -\tfrac{3}{8}(\nu-\mu)+\tfrac{1}{8}(\mu+\nu)+\tfrac{1}{2}\varsigma\\
&\qquad\qquad \le -\tfrac{1}{8}+\tfrac{3}{2}\varrho+\tfrac{5}{4}\varsigma+\tfrac{3}{2}\eta \le -\eta \quad \text{for} \quad \eta \le \tfrac{1}{20}(1-12\varrho-10\varsigma).
\end{aligned}$$

The above readily implies that validity of (10.8) subject to the conditions given in (10.9). This completes the proof of Theorem 1.2.

11. Proof of Theorem 1.3

11.1. **Automorphic preliminaries.** We present here some fundamental properties of automorphic forms, drawing primarily from [BM15], [Iwa95], and [DI82]. Let $\mathfrak{a}$ denote a cusp of $\Gamma_0(Q)$ with $\sigma_{\mathfrak{a}}$ being the corresponding scaling matrix. For a holomorphic modular form f of level Q and weight k, its Fourier expansion around $\mathfrak{a}$ is given by

$$f(\sigma_{\mathfrak{a}}z) = \sum_{n\ge1} \rho_f(\mathfrak{a},n)(4\pi n)^{\frac{k}{2}} e(nz).$$

For a Maaß form f with spectral parameter t, the expansion takes the form

$$f(\sigma_{\mathfrak{a}}z) = \sum_{n\ne0} \rho_f(\mathfrak{a},n) W_{0,it}(4\pi|n|y)e(nx),$$

where $W_{0,it}(y) = (y/\pi)^{\frac{1}{2}} K_{it}(y/2)$ is the Whittaker function. Associated to each cusp $\mathfrak{c}$ of $\Gamma_0(Q)$ is an Eisenstein series $E_{\mathfrak{c}}(\sigma_{\mathfrak{a}} z, s)$, whose Fourier expansion at $s = \frac{1}{2} + it$ is

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

For convenience, we write $\rho_f(n)$ and $\rho_{\mathfrak{c}}(n,t)$ when $\mathfrak{a} = \infty$ for the Fourier coefficients.

For a newform f, its normalized Hecke eigenvalues $\lambda(n)$ satisfy the scaling relation

$$\lambda(n)\rho_f(1) = \sqrt{n}\rho_f(n),$$

and the multiplicativity formula

$$\lambda(mn) = \sum_{d|(m,n)} \mu(d)\chi_0(d)\lambda\left(\frac{m}{d}\right)\lambda\left(\frac{n}{d}\right), \tag{11.1}$$

where χ_0 denotes the trivial character modulo Q. The Ramanujan–Petersson conjecture implies the pointwise bound

$$|\lambda(n)| \leq \tau(n),$$

which was proven for holomorphic forms in Deligne's seminal work [Del74]. In the Maaß case, the current best result is due to Kim-Sarnak [Kim03], which asserts that

$$|\lambda(n)| \leq \tau(n) n^{\theta} \quad \text{with} \quad \theta = \tfrac{7}{64}. \tag{11.2}$$

While the full conjecture remains open, its averaged version

$$\sum_{n\leq x} |\lambda(n)|^2 \ll x^{1+\varepsilon},$$

is well-established and frequently serves as a substitute for the original conjecture in applications.

For non-newforms, the exact multiplicativity (11.1) no longer holds for their Fourier coefficients, but effective averaged substitutes are available. For any complex sequence a_n, V. Blomer, G. Harcos, and P. Michel [BHM07, p. 80] established for Eisenstein series the inequality

$$\sum_{\mathfrak{c}} \left| \sum_{(n,q)=1} a_n \sqrt{qn} \rho_{\mathfrak{c}}(qn,\kappa) \right|^2 \leq 9\tau(Q)^3 \tau(q)^4 \sum_{g|(q,Q)} \sum_{\mathfrak{c}} \left| \sum_{(n,q)=1} a_n \sqrt{gn}\, \rho_{\mathfrak{c}}(gn,\kappa) \right|^2. \tag{11.3}$$

For holomorphic cusp forms of level Q and weight k, we employ the newform theory to construct special L^2-bases $\mathcal{B}_k(Q)$, comprising of forms $f|_d(z) := f^*(dz)$ where f^* is a normalized newform of level $Q_1 \mid \frac{Q}{d}$. The Fourier coefficients of forms in $\mathcal{B}_k(Q)$ exhibit approximate multiplicativity. A similar basis $\mathcal{B}(Q)$ exists for Maaß cusp forms of level Q. The explicit construction of these bases is detailed in [BHM07, Sec.3.1]. More precisely [BHM07, p.74], if $f \in \mathcal{B}_k(Q)$ (or $\mathcal{B}(Q)$) and $(s,n) = 1$, then

$$\sqrt{sn}\, \rho_f(sn) = \sum_{\delta|(\frac{s}{(s,Q)},Q)} \mu(\delta)\chi_0(\delta)\lambda_{f^*}\left(\frac{s}{\delta(s,Q)}\right)\left(\frac{(s,Q)n}{\delta}\right)^{\frac{1}{2}} \rho_f\left(\frac{(s,Q)n}{\delta}\right), \tag{11.4}$$

where f^* denotes the underlying newform. When $(n, sQ) = 1$, this simplifies to

$$\sqrt{sn}\, \rho_f(sn) = \lambda_{f^*}(n)\sqrt{s}\, \rho_f(s). \tag{11.5}$$

Furthermore [BM15, (5.4)], if f^* satisfies the Ramanujan conjecture, then

$$\left|\sum_{(n,q)=1} a_n\sqrt{qn}\rho_f(qn)\right|^2 \leq \tau(s)^2 \sum_{g|(q,Q)} \left|\sum_{(n,q)=1} a_n\sqrt{gn}\,\rho_f(gn)\right|^2.$$

Throughout this paper, we fix $\mathcal{B}_k(Q)$ and $\mathcal{B}(Q)$ to be these specially constructed bases.

11.2. **Kuznetsov formula and spectral large sieve.** Let $\mathfrak{a}$ and $\mathfrak{b}$ be two cusps of the Hecke congruence group $\Gamma_0(Q)$, and let $S_{\mathfrak{a},\mathfrak{b}}(m,n;\gamma)$ denote the associated Kloosterman sum defined in (1.9). For $Q = uv$ with $(u,v) = 1$, consider $S_{\infty,1/u}(m,n;\gamma)$ with the scaling matrix $\sigma_{1/u} = \begin{pmatrix} \sqrt{v} & 0 \\ u\sqrt{v} & \frac{1}{\sqrt{v}} \end{pmatrix}$. The Kloosterman sum $S_{\infty,1/u}(m,n;\gamma)$ is well-defined precisely when $\gamma = lu\sqrt{v}$ for some integer l coprime to v. In this case

$$S_{\infty,1/u}(m,n;\gamma) = e\left(n\frac{\overline{u}}{v}\right) S(m\overline{v},n;lu),$$

where $\overline{u}$ satisfies $u\overline{u} \equiv 1 \pmod v$ and $\overline{v}$ satisfies $v\overline{v} = 1 \pmod{lu}$ (see [DI82, formula (1.6)]).

The following lemma presents the Kuznetsov trace formula, which can be found in [Iwa95, Theorems 9.4, 9.5, 9.7].

Lemma 11.1 (Kuznetsov formula)**.** *Let $\phi(x)$ be a twice continuously differentiable function on $[0,\infty)$ satisfying $\phi(0) = 0$ and $\phi^{(j)}(x) \ll (1+x)^{-2-\varepsilon}$ for $j = 0,1,2$. Then for any cusps $\mathfrak{a}$, $\mathfrak{b}$ of $\Gamma = \Gamma_0(Q)$ and any positive integers m,n, we have*

$$\begin{aligned}\sum_{\gamma} \frac{1}{\gamma} S_{\mathfrak{a},\mathfrak{b}}(m,n;\gamma)\phi\left(\frac{4\pi\sqrt{mn}}{\gamma}\right) = &\sum_{2\leq k\equiv 0(\bmod 2)} \sum_{f\in\mathcal{B}_k(Q)} \Gamma(k)\tilde{\phi}(k)\sqrt{mn}\,\overline{\rho}_f(\mathfrak{a},m)\,\rho_f(\mathfrak{b},n)\\ &+ \sum_{f\in\mathcal{B}(Q)} \hat{\phi}(\kappa_f)\frac{\sqrt{mn}}{\cosh(\pi\kappa_f)}\,\overline{\rho}_f(\mathfrak{a},m)\,\rho_f(\mathfrak{b},n)\\ &+ \frac{1}{4\pi}\sum_{\mathfrak{c}} \int_{-\infty}^{\infty} \hat{\phi}(\kappa)\frac{\sqrt{mn}}{\cosh(\pi\kappa)}\,\overline{\rho}_{\mathfrak{a},\mathfrak{c}}(m,\kappa)\,\rho_{\mathfrak{b},\mathfrak{c}}(n,\kappa)\mathrm{d}\kappa,\end{aligned}$$

and

$$\begin{aligned}\sum_{\gamma} \frac{1}{\gamma} S_{\mathfrak{a},\mathfrak{b}}(m,-n;\gamma)\phi\left(\frac{4\pi\sqrt{mn}}{\gamma}\right) = &\sum_{f\in\mathcal{B}(Q)} \check{\phi}(\kappa_f)\frac{\sqrt{mn}}{\cosh(\pi\kappa_f)}\,\overline{\rho}_f(\mathfrak{a},m)\,\rho_f(\mathfrak{b},-n)\\ &+\frac{1}{4\pi}\sum_{\mathfrak{c}} \int_{-\infty}^{\infty} \check{\phi}(\kappa)\frac{\sqrt{mn}}{\cosh(\pi\kappa)}\,\overline{\rho}_{\mathfrak{a},\mathfrak{c}}(m,\kappa)\,\rho_{\mathfrak{b},\mathfrak{c}}(-n,\kappa)\mathrm{d}\kappa,\end{aligned}$$

where the sum is over all positive real γ for which the Kloosterman sum is well-defined. The Bessel transforms are given by

$$\tilde{\phi}(k) = 4i^k \int_0^\infty \phi(x) J_{k-1}(x) \frac{\mathrm{d}x}{x},$$
$$\hat{\phi}(\kappa) = 2\pi i \int_0^\infty \phi(x) \frac{J_{2i\kappa}(x) - J_{-2i\kappa}(x)}{\sinh(\pi\kappa)} \frac{\mathrm{d}x}{x},$$
$$\check{\phi}(\kappa) = 8 \int_0^\infty \phi(x) \cosh(\pi\kappa) K_{2i\kappa}(x) \frac{\mathrm{d}x}{x}.$$

The spectral large sieve inequality, often employed in conjunction with the Kuznetsov formula, is established as follows (see [DI82, Theorem 2]).

Lemma 11.2 (Spectral large sieve)**.** *For $K, N \geq 1$ and complex coefficients $(a_n)_{n\in[N,2N]}$, let $\mathfrak{a}$ be a cusp of $\Gamma_0(Q)$ equivalent to $\frac{u}{w}$ with $(u, v) = 1$ and $w \mid Q$. Then, for all such cusp $\mathfrak{a}$, all three quantities*

$$\sum_{\substack{2\leq k\leq K \\ k \text{ even}}} \Gamma(k) \sum_{f\in\mathcal{B}_k(Q)} \left| \sum_n a_n \sqrt{n}\, \rho_f(\mathfrak{a}, n) \right|^2, \qquad \sum_{\substack{f\in\mathcal{B}(Q) \\ |\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)} \left| \sum_n a_n \sqrt{n}\, \rho_f(\mathfrak{a}, \pm n) \right|^2,$$

$$\sum_{\mathfrak{c}} \int_{-K}^{K} \frac{1}{\cosh(\pi\kappa)} \left| \sum_n a_n \sqrt{n}\, \rho_{\mathfrak{a},\mathfrak{c}}(\pm n, \kappa) \right|^2 d\kappa,$$

are bounded by

$$(K^2 + \mu(\mathfrak{a}) N^{1+\varepsilon}) \|a_n\|_2^2,$$

where $\mu(\mathfrak{a}) = Q^{-1}$ when $\mathfrak{a} = \infty$ and $\mu(\mathfrak{a}) = \left(w, \frac{Q}{w}\right) Q^{-1}$ otherwise.

11.3. **Extension of the spectral large sieve.** To prove Theorem 1.3, we apply the Kuznetsov formula (Lemma 11.1), which leads us to consider a spectral sum of the form

$$\sum_{\substack{f\in\mathcal{B}(Q) \\ |\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)} \left| \sum_n a_n \sqrt{sn}\, \rho_f(sn) \right|^2. \tag{11.6}$$

A direct application of the spectral large sieve (Lemma 11.2) yields the bound

$$\ll (sN)^\varepsilon \left(K^2 + \frac{sN}{Q} \right) \|a_n\|_2^2,$$

where the dependence on s introduces a significant loss, especially when sN becomes large. Our main goal is to derive an improved bound that effectively mitigates the impact of the parameter s.

For both the holomorphic and the Eisenstein cases, the Ramanujan–Petersson conjecture is well-established. Consequently, we can obtain bounds for (11.6) by applying Lemma 11.2, leveraging the almost multiplicativity property of the Fourier coefficients and the Ramanujan–Petersson conjecture.

Theorem 11.3. *For $Q, s \in \mathbb{N}$, $K, N \geq 1$ and complex coefficients $(a_n)_{n\in[N,2N]}$, both two quantities*

$$\sum_{\substack{2\leq k\leq K\\ k\ even}} \Gamma(k) \sum_{f\in\mathcal{B}_k(Q)} \left|\sum_n a_n\sqrt{sn}\,\rho_f(sn)\right|^2, \quad \sum_{\mathfrak{c}} \int_{-K}^{K} \frac{1}{\cosh(\pi\kappa)} \left|\sum_n a_n\sqrt{sn}\,\rho_{\mathfrak{c}}(\pm sn,\kappa)\right|^2 \mathrm{d}\kappa$$

are bounded by

$$(sQN)^{\varepsilon}\left(K^2 + \frac{(s,Q)N}{Q}\right)N\|a_n\|_\infty. \tag{11.7}$$

Proof. We first treat the holomorphic case, with the Eisenstein spectrum proceeding analogously. By performing the substitution $n \to dn'$ where $d = (n, s^\infty)$, we decompose the sum as

$$\sum_n a_n\sqrt{sn}\rho_f(sn) \ll \sum_{d|s^\infty}\left|\sum_{(n',ds)=1} a_{dn'}\sqrt{dsn'}\,\rho_f(dsn')\right|.$$

Applying the multiplicativity relation (11.4) with $g = (ds, Q)/\delta$ and invoking Deligne's bound yields

$$\sum_{(n',ds)=1} a_{dn'}\sqrt{dsn'}\,\rho_f(dsn') \ll (sN)^{\varepsilon} \sum_{g|(ds,Q)}\left|\sum_{(n',ds)=1} a_{dn'}\sqrt{gn'}\,\rho_f(gn')\right|.$$

Employing the Cauchy-Schwarz inequality over the sum in d and g, we deduce that

$$\left|\sum_n a_n\sqrt{sn}\rho_f(sn)\right|^2 \ll (sQN)^{\varepsilon}\sum_{d|s^\infty}\sum_{g|(ds,Q)}\left|\sum_{(n',s)=1} a_{dn'}\sqrt{gn'}\,\rho_f(gn')\right|^2,$$

where the factor $(sQN)^{\varepsilon}$ accounts for the number of terms d and g. Inserting this into the holomorphic spectrum and applying Lemma 11.2 with $n = gn' \leq (s,Q)N$ establishes the bound (11.7). The Eisenstein case follows identically, replacing (11.4) with the corresponding estimate (11.3). □

For the Maaß case in which the Ramanujan–Petersson conjecture remains open, additional techniques are necessary to manage the s-dependence. When $a_n \leq 1$, V. Blomer and D. Milićević [BM15, Theorem 13] established the fundamental bound

$$\sum_{\substack{f\in\mathcal{B}(Q)\\ |\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)}\left|\sum_{(n,sQ)=1} a_n\sqrt{sn}\,\rho_f(sn)\right|^2 \ll (sQKN)^{\varepsilon}(s,Q)\left(K + \frac{s^{\frac12}}{Q^{\frac12}}\right)\left(K + \frac{N}{Q^{\frac12}}\right)N. \tag{11.8}$$

While this bound requires the sum over $(n, sQ) = 1$ and depends on s through a factor (s, Q) that is typically small (e.g., for cusp forms on $SL_2(Z)$), it can become substantial large for the Hecke congruence groups of high level, potentially yielding weaker estimates than applying the Kim-Sarnak upper bound for the Hecke eigenvalue. Note that this factor (s, Q) arises from the coefficient estimate

$$\sqrt{n}\,\rho_f(n) \ll (nQ)^{\varepsilon} n^{\theta}(Q,n)^{\frac12-\theta}|\,\rho_{f^*}(1)|.$$

To overcome this limitation, we develop a new relationship in the following.

Proposition 11.4. *For $f \in \mathcal{B}(Q)$ or $\mathcal{B}_k(Q)$ with associated newform f^*, let $s, s' \in \mathbb{N}$ with $s' \mid \frac{s}{(s,Q)}$. Then we have*

$$\sqrt{s}\,\rho_f(s)\lambda_{f^*}(s') = \sum_{g \mid s'} \left(\frac{ss'}{g^2}\right)^{\frac{1}{2}} \rho_f\left(\frac{ss'}{g^2}\right).$$

Proof. By combining (11.4) with the Möbius inversion of (11.1), we obtain the identity

$$\begin{aligned}\sqrt{s}\,\rho_f(s)\lambda_{f^*}(s') &= \sum_{\delta \mid \left(\frac{s}{(s,Q)},Q\right)} \mu(\delta)\chi_0(\delta)\lambda_{f^*}\left(\frac{s}{\delta(s,Q)}\right)\lambda_{f^*}(s')\left(\frac{(s,Q)}{\delta}\right)^{\frac{1}{2}} \rho_f\left(\frac{(s,Q)}{\delta}\right)\\
&= \sum_{g \mid s'} \sum_{\delta \mid \left(\frac{s}{g(s,Q)},Q\right)} \mu(\delta)\chi_0(\delta)\lambda_{f^*}\left(\frac{ss'}{g^2\delta(s,Q)}\right)\left(\frac{(s,Q)}{\delta}\right)^{\frac{1}{2}} \rho_f\left(\frac{(s,Q)}{\delta}\right). \qquad (11.9)\end{aligned}$$

Setting $s'' = \frac{ss'}{g^2}$ for $g \mid s' \mid \frac{s}{(s,Q)}$, we establish the relations

$$(s,Q) = ((s,Q),Q) \le \left(\frac{s}{g},Q\right) \le \left(\frac{ss'}{g^2},Q\right) = (s'',Q),$$

$$(s'',Q) = \left(\frac{ss'}{g^2},Q\right) \le (ss',Q) \le \left(\frac{s^2}{(s,Q)},Q\right) = (s,Q).$$

These indicate that $(s,Q) = (s'',Q)$, yielding

$$\left(\frac{ss'}{g^2(s,Q)},Q\right) = \left(\frac{s''}{(s'',Q)},Q\right),$$

while

$$\left(\frac{s}{g(s,Q)},Q\right) \mid \left(\frac{ss'}{g^2(s,Q)},Q\right) \mid \left(\frac{s^2}{g^2(s,Q)^2},Q\right).$$

These relations indicate that $\left(\frac{s}{g(s,Q)},Q\right)$ and $\left(\frac{s''}{(s'',Q)},Q\right)$ share the same distinct prime factors. This identical prime factor structure allows us to simplify the δ-sum in (11.9) via (11.4), resulting in

$$\sum_{\delta \mid \left(\frac{s''}{(s'',Q)},Q\right)} \mu(\delta)\chi_0(\delta)\lambda_{f^*}\left(\frac{s''}{\delta(s'',Q)}\right)\left(\frac{(s'',Q)}{\delta}\right)^{\frac{1}{2}} \rho_f\left(\frac{(s'',Q)}{\delta}\right) = \sqrt{s''}\,\rho_f(s''),$$

thereby completing the proof. □

We also need the following estimate [BM15, Lemma 12], which is another application of the Kuznetsov formula.

Lemma 11.5. *For $K \ge 1$, $Q, n \in \mathbb{N}$, we have*

$$\sum_{\substack{f \in \mathcal{B}(Q)\\ |\kappa_f| \le K}} \frac{1}{\cosh(\pi\kappa_f)} \left|\sqrt{n}\,\rho_f(n)\right|^2 \ll (Kn)^{\varepsilon}\left(K^2 + \frac{(Q,n)^{\frac{1}{2}} n^{\frac{1}{2}}}{Q}\right).$$

Building upon Proposition 11.4 and employing a refined application of the Cauchy-Schwarz inequality, we achieve two significant improvements. First, we substantially eliminate the restrictive (s, Q) factor presented in (11.8). Second, we completely relax the previously essential coprimality condition $(n, sQ) = 1$. This relaxation is particularly advantageous when analyzing the parameters h and k in Theorems 1.1–1.2, where the level Q becomes dependent on k and h. The resulting bound, formulated in the following theorem, demonstrates both remarkable generality and simplicity, rendering it widely applicable.

Theorem 11.6. *Let $Q, s \in \mathbb{N}$, $K, N \geq 1$, and let $(a_n)_{n\in[N,2N]}$ be complex coefficients. Then we have*

$$\sum_{\substack{f\in\mathcal{B}(Q)\\|\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)}\left|\sum_n a_n\sqrt{sn}\,\rho_f(sn)\right|^2 \ll (sQKN)^{\varepsilon}\left(K+\frac{s^{\frac12}}{Q^{\frac12}}\right)\left(K+\frac{(s,Q)^{\frac12}N}{Q^{\frac12}}\right)N\|a_n\|_\infty.$$

Proof. We begin by analyzing the case where $(n, sQ) = 1$, defining

$$\mathcal{G}(s,N) = \sum_{\substack{f\in\mathcal{B}(Q)\\|\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)}\left|\sum_{(n,sQ)=1} a_n\sqrt{sn}\,\rho_f(sn)\right|^2.$$

We first utilize the almost multiplicativity from (11.5) to separate the variables s and n and then we apply (11.4) to $\sqrt{s}\rho_f(s)$, obtaining the decomposition

$$\sqrt{s}\,\rho_f(s) = \sum_{\delta|\left(\frac{s}{(s,Q)},Q\right)} \mu(\delta)\chi_0(\delta)\lambda_{f^*}\left(\frac{s}{\delta(s,Q)}\right)\left(\frac{(s,Q)}{\delta}\right)^{\frac12}\rho_f\left(\frac{(s,Q)}{\delta}\right).$$

This leads to the inequality

$$\begin{aligned}\mathcal{G}(s,N) \leq \sum_{\delta|\left(\frac{s}{(s,Q)},Q\right)} \sum_{\substack{f\in\mathcal{B}(Q)\\|\kappa_f|\leq K}} &\frac{1}{\cosh(\pi\kappa_f)}\left|\sum_{(n,sQ)=1} a_n\lambda_{f^*}(n)\right|^2\\ &\times\left|\sqrt{s}\,\rho_f(s)\lambda_{f^*}\left(\frac{s}{\delta(s,Q)}\right)\left(\frac{(s,Q)}{\delta}\right)^{\frac12}\rho_f\left(\frac{(s,Q)}{\delta}\right)\right|.\end{aligned}$$

Through application of the Cauchy-Schwarz inequality to the spectral sum, we derive

$$\mathcal{G}(s,N) \leq \sum_{\delta|\left(\frac{s}{(s,Q)},Q\right)} \mathcal{G}_1^{\frac12}\mathcal{G}_2^{\frac12},$$

where the components are defined as

$$\mathcal{G}_1 = \sum_{\substack{f\in\mathcal{B}(Q)\\|\kappa_f|\leq K}} \frac{1}{\cosh(\pi\kappa_f)}\left|\sqrt{s}\,\rho_f(s)\lambda_{f^*}\left(\frac{s}{\delta(s,Q)}\right)\right|^2,$$

and

$$\mathcal{G}_2 = \sum_{\substack{f\in\mathcal{B}(Q)\\ |\kappa_f|\le K}} \frac{1}{\cosh(\pi\kappa_f)}\Big|\sum_{(n,sQ)=1} a_n\lambda_{f^*}(n)\Big|^4\Big|\Big(\frac{(s,Q)}{\delta}\Big)^{\frac12}\rho_f\Big(\frac{(s,Q)}{\delta}\Big)\Big|^2.$$

Proposition 11.4 applied with $s'=\frac{s}{\delta(s,Q)}$ yields

$$\mathcal{G}_1 \ll s^{\varepsilon}\sum_{g|s'}\sum_{\substack{f\in\mathcal{B}(Q)\\ |\kappa_f|\le K}} \frac{1}{\cosh(\pi\kappa_f)}\Big|\Big(\frac{ss'}{g^2}\Big)^{\frac12}\rho_f\Big(\frac{ss'}{g^2}\Big)\Big|^2.$$

From Lemma 11.5 and the identity $\left(\frac{ss'}{g^2},Q\right)=(s,Q)$ established in Proposition 11.4, we obtain the bound

$$\mathcal{G}_1 \ll (sK)^{\varepsilon}\Big(K^2+\frac{s}{Q}\Big).$$

For $\mathcal{G}_2$, employing the Möbius inversion in (11.1) together with (11.5) gives

$$\begin{aligned}\mathcal{G}_2 &= \sum_{\substack{f\in\mathcal{B}(Q)\\ |\kappa_f|\le K}} \frac{1}{\cosh(\pi\kappa_f)}\Big|\sum_{(n_1n_2,sQ)=1} a_{n_1}a_{n_2}\sum_{r|(n_1,n_2)}\lambda_{f^*}\Big(\frac{n_1n_2}{r^2}\Big)\Big(\frac{(s,Q)}{\delta}\Big)^{\frac12}\rho_f\Big(\frac{(s,Q)}{\delta}\Big)\Big|^2\\ &\ll \sum_{\substack{f\in\mathcal{B}(Q)\\ |\kappa_f|\le K}} \frac{1}{\cosh(\pi\kappa_f)}\Big|\sum_{n\le N^2} b_n\Big(n\frac{(s,Q)}{\delta}\Big)^{\frac12}\rho_f\Big(n\frac{(s,Q)}{\delta}\Big)\Big|^2,\end{aligned}$$

where the coefficients satisfy

$$b_n = \sum_{(n_1n_2,sQ)=1} a_{n_1}a_{n_2}\sum_{\substack{r|(n_1,n_2)\\ n_1n_2=r^2n}} 1 \ll \|a_n\|_\infty^2\sum_{r\ll N/\sqrt{n}}\tau(n) \ll \frac{N^{1+\varepsilon}}{n^{\frac12}}\|a_n\|_\infty^2.$$

Applying the spectral large sieve (Lemma 11.2) reveals that

$$\mathcal{G}_2 \ll N^{\varepsilon}\Big(K^2+\frac{(s,Q)N^2}{Q}\Big)N^2\|a_n\|_\infty^2.$$

Combining these estimates, we establish for $(n,sQ)=1$ the fundamental bound

$$\mathcal{G}(s,N) \ll (sKN)^{\varepsilon}\Big(K+\frac{s^{\frac12}}{Q^{\frac12}}\Big)\Big(K+\frac{(s,Q)^{\frac12}N}{Q^{\frac12}}\Big)N\|a_n\|_\infty. \tag{11.10}$$

For the general case without $(n,sQ)=1$, we employ the following decomposition

$$\sum_n a_n\sqrt{sn}\,\rho_f(sn) = \sum_{d|(sQ)^\infty}\sum_{(n,sdQ)=1} a_{dn}\sqrt{sdn}\,\rho_f(sdn).$$

Given that the d-sum contains $\ll (sQN)^\varepsilon$ terms, an application of the Cauchy-Schwarz inequality yields

$$\sum_{\substack{f\in\mathcal{B}(Q)\\|\kappa_f|\le K}} \frac{1}{\cosh(\pi\kappa_f)}\Big|\sum_n a_n\sqrt{sn}\,\rho_f(sn)\Big|^2 \ll (sQN)^\varepsilon \sum_{\substack{d\le 2N\\ d|(sQ)^\infty}} \mathcal{G}\Big(sd,\frac{N}{d}\Big)$$
$$\ll (sQKN)^\varepsilon\Big(K+\frac{s^{\frac12}}{Q^{\frac12}}\Big)\Big(K+\frac{(s,Q)^{\frac12}N}{Q^{\frac12}}\Big)N\|a_n\|_\infty,$$

where we have invoked the bound (11.10).[1] This completes the proof. □

11.4. **Completion of the proof.** We want to apply the Kuznetsov's formula (Lemma 11.1) to treat the left-hand side of (1.13). Before doing so, we review key estimates for the Bessel transforms from Lemma 11.1. As established in [DI82, Lemma 7.1], the following bounds hold for real κ.

$$\hat{\phi}(i\kappa),\ \check{\phi}(i\kappa) \ll \frac{1+X^{-2\kappa}}{1+X} \quad \text{for } 0<\kappa<\tfrac12, \tag{11.11a}$$

$$\tilde{\phi}(\kappa),\ \hat{\phi}(\kappa),\ \check{\phi}(\kappa) \ll \frac{1+|\log X|}{1+X} \quad \text{for all } \kappa, \tag{11.11b}$$

$$\tilde{\phi}(\kappa),\ \hat{\phi}(\kappa),\ \check{\phi}(\kappa) \ll \Big(1+\kappa^{-\frac12}X\Big)|\kappa|^{-\frac52} \quad \text{for all } |\kappa|\ge 1. \tag{11.11c}$$

We now focus on the situation where m and n in the Kloosterman sum share the same sign (the complementary case can be treated analogously). Applying Lemma 11.1 with $\mathfrak{a}=\infty$ and $\mathfrak{b}=1/u$, we bound the left-hand side of (1.13) by decomposing it into three principal contributions corresponding to the spectral decomposition

$$\begin{aligned}
\ll &\sum_{2\le k\equiv 0(\bmod 2)}\ \sum_{f\in\mathcal{B}_k(Q)} \Gamma(k)|\tilde{\phi}(k)|\Big|\sum_m a_m\sqrt{sm}\,\overline{\rho}_f(sm)\Big|\Big|\sum_n b_n\sqrt{n}\,\rho_f(\mathfrak{b},n)\Big|\\
&+\sum_{f\in\mathcal{B}(Q)} \frac{|\hat{\phi}(\kappa_f)|}{\cosh(\pi\kappa_f)}\Big|\sum_m a_m\sqrt{sm}\,\overline{\rho}_f(sm)\Big|\Big|\sum_n b_n\sqrt{n}\,\rho_f(\mathfrak{b},n)\Big|\\
&+\frac{1}{4\pi}\sum_{\mathfrak{c}}\int_{-\infty}^{\infty}\frac{|\hat{\phi}(\kappa)|}{\cosh(\pi\kappa)}\Big|\sum_m a_m\sqrt{sm}\,\overline{\rho}_{\mathfrak{c}}(sm,\kappa)\Big|\Big|\sum_n b_n\sqrt{n}\,\rho_{\mathfrak{b},\mathfrak{c}}(n,\kappa)\Big|\mathrm{d}\kappa.
\end{aligned}$$

These terms respectively account for the holomorphic, Maaß, and continuous spectrum contributions. We now focus our analysis on the Maaß spectrum (the other cases follow similarly), decomposing it as

$$\Sigma_1+\sum_{K\gg 1+X}\Sigma_K,$$

where Σ_1 aggregates terms with $\kappa_f \ll 1+X$ and Σ_K sums over $\kappa_f \asymp K$ for $K\gg 1+X$. After applying the Cauchy-Schwarz inequality, we estimate the n-sum via Lemma 11.2

[1] Note that employing estimate (11.8) directly would introduce an additional factor $Q^{\frac12}$ in the final bound.

and the m sum through Theorem 11.6, yielding

$$\Sigma_1 \ll \frac{1+|\log X|+X^{-\vartheta_Q}}{1+X}\left(1+X+\frac{s^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1+X+\frac{(s,Q)^{\frac{1}{2}}M}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(1+X+\frac{N^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)M^{\frac{1}{2}}\|a_m\|_\infty^{\frac{1}{2}}\|b_n\|_2,$$

where we employ (11.11a) and (11.11b) to bound the Bessel transforms, and

$$\Sigma_K \ll \left(1+K^{-\frac{1}{2}}X\right)K^{-\frac{5}{2}}\left(K+\frac{s^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(K+\frac{(s,Q)^{\frac{1}{2}}M}{Q^{\frac{1}{2}}}\right)^{\frac{1}{2}}\left(K+\frac{N^{\frac{1}{2}}}{Q^{\frac{1}{2}}}\right)M^{\frac{1}{2}}\|a_m\|_\infty^{\frac{1}{2}}\|b_n\|_2$$

via (11.11c). Summing over K recovers the bound in (1.13). The holomorphic and continuous spectrum are treated similarly, with Theorem 11.6 being replaced by Theorem 11.3 wherever appropriate. This completes the proof of Theorem 1.3.

Acknowledgments. P. Gao is supported in part by the NSFC grant 12471003, X. Wu by the NSFC Grant 12271135, Anhui Provincial Natural Science Foundation Grant 2508085J005 and the Fundamental Research Funds for the Central Universities Grant JZ2025HGTG0254, and L. Zhao by the FRG Grant PS71536 at the University of New South Wales.

References

[BBLR16] S. Bettin, H. M. Bui, X. Li, and M. Radziwiłł, A quadratic divisor problem and moments of the Riemann zeta-function, J. Eur. Math. Soc. (JEMS) 22 (2020), 3953–3980. 4, 9

[BFKMM17+a] V. Blomer, É. Fouvry, E. Kowalski, Ph. Michel, and D. Milićević, On moments of twisted L-functions, Amer. J. Math. 139 (2017), 707–768. 2, 4

[BFKMM17+b] V. Blomer, É. Fouvry, E. Kowalski, Ph. Michel, and D. Milićević, Some applications of smooth bilinear forms with Kloosterman sums, Tr. Mat. Inst. Steklova 296 (2017), 24–35 (Russian). English transl. in Proc. Steklov Inst. Math. 296 (2017), 18–29. 2, 4

[BFKMMS23] V. Blomer, É. Fouvry, E. Kowalski, P. Michel, D. Milićević, and W. Sawin, The second moment theory of families of L-functions—the case of twisted Hecke L-functions, Mem. Amer. Math. Soc. 282(2023), v+148. 2, 4

[BHKM20] V. Blomer, P. Humphries, R. Khan, and M. B. Milinovich, Motohashi's fourth moment identity for non-archimedean test functions and applications, Compos. Math. 156 (2020), 1004–1038. 4, 5

[BHM07] V. Blomer, G. Harcos, and P. Michel, A Burgess-like subconvex bound for twisted L-functions (with appendix 2 by Z. Mao), Forum Math. 19 (2007), 61–105. 33

[BM15] V. Blomer and D. Milićević, The second moment of twisted modular L-functions, Geom. Funct. Anal. 25 (2015), 453–516. 6, 32, 34, 36, 37

[BPRZ20] H. M. Bui, K. Pratt, N. Robles, and A. Zaharescu, Breaking the $\frac{1}{2}$-barrier for the twisted second moment of Dirichlet L-functions, Adv. Math. 370 (2020), 107175, 40pp.

[CFKRS05] J. B. Conrey, D. W. Farmer, J. P. Keating, M. O. Rubinstein, and N. C. Snaith, Integral moments of L-functions, Proc. London Math. Soc. (3) 91 (2005), 33–104. 1

[CS07] J. B. Conrey and N. C. Snaith, Applications of the L-functions ratios conjectures, Proc. Lond. Math. Soc. (3) 94 (2007), 594–646 2

[Del74] P. Deligne, La conjecure de Weil, I, Inst. Hautes Études Sci. Publ. Math. 43 (1974), 273–307. 33

[DI82] J. M. Deshouillers and H. Iwaniec, Kloosterman sums and fourier coefficients of cusp forms, Invent. Math. 70 (1982), 219–288. 6, 7, 32, 34, 35, 40

[GZ25+a] P. Gao and L. Zhao, Upper bounds for moments of Dirichlet L-functions to a fixed modulus, Mathematika, to appear. 5

[GZ25+b] P. Gao and L. Zhao, Twisted fourth moment of Dirichlet L-functions to a fixed modulus, arXiv:2507.18186. 2, 4

[HB81] D. R. Heath-Brown, The fourth power mean of Dirichlet's L-functions, Analysis 1 (1981), 25–32. 2, 8

[Hou16] B. Hough, The angle of large values of L-functions, J. Number Theory 167 (2016), 353–393. 2, 4, 5

[HM10] C. P. Hughes and P. Y. Matthew, The twisted fourth moment of the Riemann zeta function. J. Reine Angew. Math 641 (2010), 203–236. 4

[Iwa95] H. Iwaniec, Introduction to the Spectral Theory of Automorphic Forms, Bibl. Rev. Mat. Iberoamericana, Revista Matemática Iberoamericana, Madrid (1995). 32, 34

[Kim03] H. H. Kim, Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2, J. Amer. Math. Soc. 16 (2003), no. 1, 139–183, with Appendix 1 by Dinakar Ramakrishnan and Appendix 2 by Kim and Peter Sarnak. 33

[KSWX23] B. Kerr, I. Shparlinski, X. Wu, and P. Xi, Bounds on bilinear forms with Kloosterman sums, J. London Math. Soc. (2) 108 (2023), 578–621. 5, 13, 29

[Liu24] D. Liu, Zeros and moments of L-functions and applications, Thesis (Ph.D.)–University of Illinois Urbana-Champaign (2024), 95pp. 2, 4

[Mun17] M. Munsch, Shifted moments of L-functions and moments of theta functions, Mathematika 63 (2017), 196–212. 2

[RS15] M. Radziwiłł and K. Soundararajan, Moments and distribution of central L-values of quadratic twists of elliptic curves, Invent. Math. 202 (2015), 1029–1068. 2

[RS05] Z. Rudnick and K. Soundararajan, Lower bounds for moments of L-functions, Proc. Natl. Acad. Sci. USA, 102 (2005), 6837–6838. 1

[Sel46] A. Selberg, Contributions to the theory of Dirichlet's L-functions, Skr. Norske Vid.-Akad. Oslo I (1946), 62pp. 2

[Sou07] K. Soundararajan, The fourth moment of Dirichlet L-functions, in Analytic Number Theory, Clay Math. Proc. 7, Amer. Math. Soc., Providence, RI, 2007, pp. 239–246. 2, 8

[Sou08] K. Soundararajan, Extreme values of zeta and L-functions, Math. Ann. 342 (2008), 467–486. 2

[Sza24] B. Szabó, High moments of theta functions and character sums, Mathematika 70, (2024), Paper No. e12242, 37 pp. 2

[Wu19] X. Wu, The twisted mean square and critical zeros of Dirichlet L-functions, Math. Z. 293 (2019), 825–865. 9

[Wu23+a] X. Wu, The fourth moment of Dirichlet L-functions at the central value, Math. Ann. 387 (2023), 1199–1248. 2, 4, 5, 9, 17

[Wu23+b] X. Wu, The fourth moment of Dirichlet L-functions along the critical line, Forum Math. 35 (2023), 1347–1371. 9

[You11] M. P. Young, The fourth moment of Dirichlet L-functions, Ann. of Math. 173 (2011), 1–50. 2, 5, 8, 26, 27

[Zac19] R. Zacharias, Mollification of the fourth moment of Dirichlet L-functions, Acta Arith. 191 (2019), 201–257. 2, 4, 5

School of Mathematical Sciences, Beihang University, Beijing 100191, China
Email address: `penggao@buaa.edu.cn`

School of Mathematics, Hefei University of Technology, Hefei 230009, China
Email address: `xswu@amss.ac.cn`

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia
Email address: `l.zhao@unsw.edu.au`