The Weyl bound for Dirichlet L-functions of cube-free conductor

By Ian Petrow and Matthew P. Young

Abstract

We prove a Weyl-exponent subconvex bound for any Dirichlet L-function of cube-free conductor. We also show a bound of the same strength for certain L-functions of self-dual GL_2 automorphic forms that arise as twists of forms of smaller conductor.

1. Introduction

Subconvex estimates for L-functions play a major role in modern analytic number theory. The first subconvex estimate is due to Weyl and Hardy-Littlewood, who showed that

$$(1.1) \qquad \zeta(1/2 + it) \ll_\varepsilon (1 + |t|)^{\frac{1}{6}+\varepsilon}.$$

The exponent $1/6$ appearing in (1.1) is a consequence of Weyl's differencing method for estimating exponential sums, introduced in 1916. This method itself is important for studying equidistribution and has immediate applications to lattice point counting problems.

Today we call a subconvex bound of the form $L(1/2, \pi) \ll_\varepsilon Q(\pi)^{1/6+\varepsilon}$ *the Weyl bound*, where $Q(\pi)$ is the analytic conductor of the automorphic L-function $L(1/2, \pi)$. The Weyl bound is only known in a few cases, notably for quadratic twists of certain self-dual GL_2 automorphic forms; see [CI00], [Ivi01], [You17], [PY19b] for example.

Estimating the Dirichlet L-functions $L(1/2, \chi)$ of conductor q as $q \to \infty$ is analogous to estimating $\zeta(1/2+it)$ as $t \to \infty$, but the former is a harder and more arithmetic problem. In 1963, Burgess [Bur63] showed by a completely different method that

$$(1.2) \qquad\qquad L(1/2, \chi) \ll_\varepsilon q^{\frac{3}{16}+\varepsilon}.$$

Burgess's method required new ideas; in particular, it uses the Riemann Hypothesis for curves over finite fields. Note that the Burgess exponent of $3/16$ falls short of the exponent $1/6$ found by Weyl. Curiously, the exponent $3/16$ often re-occurs in the modern incarnations of these problems; see [BHM07], [BH08], [BH14], [Wu14], [Wu19] for example.

Even for the case of Dirichlet L-functions, the Burgess bound has only been improved in some limited special cases. In a breakthrough, Conrey and Iwaniec [CI00] obtained a Weyl-quality bound for *quadratic* characters of odd conductor using techniques from automorphic forms and Deligne's solution of the Weil conjectures for varieties over finite fields. Another class of results, such as [BLT64] and [HB78], consider situations where the conductor q of χ runs over prime powers or otherwise has some special factorizations. Notably, Milićević [Mil16] recently obtained a sub-Weyl subconvex bound when $q = p^n$ with n large.

One of the main results of this paper (see Corollary 1.3) gives a Weyl-exponent subconvex bound for any Dirichlet L-function of cube-free conductor. In particular, we give the first improvement on the Burgess bound for all Dirichlet L-functions of prime conductor.

1.1. *Statement of results.* Let q be a positive integer, and let χ be a primitive Dirichlet character of conductor q. Let $\mathcal{H}_{it_j}(m, \overline{\chi}^2)$ denote the set (possibly empty) of Hecke-normalized Hecke-Maass newforms of level $m|q$, central character $\overline{\chi}^2$ and spectral parameter t_j. For $f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)$, $f \otimes \chi$ is a self-dual newform of level q^2 and trivial central character.

THEOREM 1.1. *Let notation be as above. Assume q is cube-free and χ is not quadratic. Then for some $B > 2$, we have*

$$(1.3) \quad \sum_{m|q} \sum_{|t_j| \leq T} \sum_{f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)} L(1/2, f \otimes \chi)^3 + \int_{-T}^{T} |L(1/2+it, \chi)|^6 dt \ll_\varepsilon T^B q^{1+\varepsilon}.$$

Theorem 1.1 generalizes the celebrated result of Conrey and Iwaniec [CI00] that assumed χ is the quadratic character of odd, square-free conductor q. The central values appearing in Theorem 1.1 are non-negative [Wal85], [Guo96], which is crucial for obtaining the Weyl-quality subconvex bound for these central values.

A potential defect of Theorem 1.1 is that, although it is consistent with the Lindelöf hypothesis in the q-aspect, it is weak in the T-aspect. However, if $T \ll q^\varepsilon$, then it is sharp. As in the work of [You17], we can obtain a hybrid result for $T \gg q^\varepsilon$.

THEOREM 1.2. *Let conditions be as in Theorem* 1.1, *and suppose that* $T \gg q^\delta$ *for some* $\delta > 0$. *Then*

(1.4)
$$\sum_{m|q} \sum_{T \le t_j < T+1} \sum_{f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)} L(1/2, f \otimes \chi)^3$$
$$+ \int_T^{T+1} |L(1/2 + it, \chi)|^6 dt \ll_{\varepsilon, \delta} T^{1+\varepsilon} q^{1+\varepsilon}.$$

As a consequence, we obtain a Weyl-quality subconvex bound for Dirichlet L-functions simultaneously in q- and t-aspects:

COROLLARY 1.3. *Suppose* χ *has cubefree conductor* q. *Then*

(1.5)
$$|L(1/2 + it, \chi)| \ll_\varepsilon q^{1/6+\varepsilon}(1 + |t|)^{1/6+\varepsilon}.$$

COROLLARY 1.4. *Let* p *be an odd prime, and suppose* F *is a Hecke-Maass newform of level* p^2, *trivial central character, and spectral parameter* t_F. *If* F *is not twist-minimal, then*

(1.6)
$$L(1/2, F) \ll_\varepsilon (p(1 + |t_F|))^{1/3+\varepsilon}.$$

Here the assumption that F is not twist minimal means there exists a newform f of level m dividing p and a primitive Dirichlet character χ of conductor p so that $F = f \otimes \chi$. The central character of F, which is trivial by assumption, equals χ^2 times the central character of f. Hence $f \in \mathcal{H}_{it_F}(m, \overline{\chi}^2)$, and so Theorem 1.1 applies. Another observation is that for F of level p^2 and trivial central character, the condition that F is twist-minimal is equivalent to the assertion that the local representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ associated to F is supercuspidal (see, e.g., [Gel75, Table (4.20)]).

Theorems 1.1 and 1.2 (and hence Corollary 1.4) also carry over to holomorphic modular forms. Let $S_\kappa(q, \overline{\chi}^2)$ denote the space of cusp forms of level q, central character $\overline{\chi}^2$, and even weight $\kappa \ge 2$. Let $\mathcal{H}_\kappa(m, \overline{\chi}^2)$ denote the set of Hecke-normalized newforms of level $m|q$ and central character $\overline{\chi}^2$.

THEOREM 1.5. *Let notation be as above, with* q *cube-free. Then*

(1.7)
$$\sum_{m|q} \sum_{\kappa \le T} \sum_{f \in \mathcal{H}_\kappa(m, \overline{\chi}^2)} L(1/2, f \otimes \chi)^3 \ll_\varepsilon T^B q^{1+\varepsilon}$$

for some $B > 2$. Moreover, if there exists $\delta > 0$ such that $T \gg q^\delta$, then we have

$$(1.8) \qquad \sum_{m|q} \sum_{T \leq \kappa < T+1} \sum_{f \in \mathcal{H}_\kappa(m, \overline{\chi}^2)} L(1/2, f \otimes \chi)^3 \ll_{\varepsilon, \delta} T^{1+\varepsilon} q^{1+\varepsilon}.$$

The sum over κ in (1.8) has at most one non-zero term and is often empty. Nonetheless, we include it so that (1.8) aligns with the form of (1.4).

1.2. *Remarks.* The reader may wonder why q is restricted to be cube-free in the above results (Coincidentally, the Burgess bound for character sums is stronger in certain ranges in case the conductor is cube-free; see, e.g., [IK04, Thm. 12.6].) To explain this restriction on q, we need to outline the proof of Theorem 1.1. As in the work of Conrey and Iwaniec [CI00], we apply some standard tools: approximate functional equations, the Petersson/Kuznetsov formula, and Poisson summation. The dual sum after Poisson summation in large part boils down to a certain character sum defined by

$$(1.9) \qquad g(\chi, \psi) = \sum_{t,u \,(\mathrm{mod}\ q)} \chi(t)\overline{\chi}(t+1)\overline{\chi}(u)\chi(u+1)\psi(ut-1),$$

where ψ is a Dirichlet character modulo q. After the above steps, the problem essentially reduces to bounding

$$(1.10) \qquad \sum_{\psi \,(\mathrm{mod}\ q)} |L(1/2, \psi)|^4 g(\chi, \psi).$$

Since the fourth moment of Dirichlet L-functions is of size $O_\varepsilon(q^{1+\varepsilon})$, the sum (1.10) can be bounded by $O_\varepsilon(q^{1+\varepsilon})$ times the maximum value of $|g(\chi, \psi)|$ as ψ varies. Here, the Riemann hypothesis of Deligne [Del80] plays a crucial role in proving $|g(\chi, \psi)| \ll_\varepsilon q^{1+\varepsilon}$ for q prime (see Section 9.1), which then extends to square-free q by multiplicativity. In case $q = p^2$, we establish $|g(\chi, \psi)| \ll_\varepsilon q^{1+\varepsilon}$ by elementary means (see Section 9.2), and hence this bound on $g(\chi, \psi)$ holds for cube-free q. However, for $q = p^3$, it is no longer true that $|g(\chi, \psi)| \ll_\varepsilon q^{1+\varepsilon}$ for all primitive ψ. Rather, there exist many characters of conductor p^3 so that $|g(\chi, \psi)| \gg q p^{1/2}$. Barring an improved estimate for the sub-sum of (1.10) coming from these "bad" characters ψ, this extra factor of $p^{1/2}$ would propagate through all the estimates, and hence would presumably lead to (at best) the bound

$$(1.11) \qquad |L(1/2 + it, \chi)|^6 \ll_\varepsilon q^{1+\varepsilon} p^{1/2} \qquad (q = p^3).$$

This would imply $|L(1/2 + it, \chi)| \ll_\varepsilon q^{\frac{7}{36}+\varepsilon}$, and note $\frac{7}{36} > \frac{3}{16}$, so this would not improve on the Burgess bound.

The analysis of $g(\chi, \psi)$ becomes more complicated for $q = p^n$ with larger n. Since there are complementary methods well-suited to treat the depth-aspect

(as in [Mil16], [BM15], and other papers), we content ourselves here with the restriction to q cube-free.

Remark added August 28, 2019: In [PY19a], written after the first version of the present paper, we have extended all the cubic moment bounds stated in Section 1.1 to hold for arbitrary q. More precisely, [PY19a] contains proofs of Conjectures 6.6 and 8.2 from the present paper, which are shown here to imply the cubic moment bounds for general q.

1.3. *Organization of the paper.* For the rest of the paper, we will focus almost entirely on the proof of Theorem 1.1. The proof of Theorem 1.2 follows the same approach, and the only change is in the behavior of the weight function on the spectrum. These archimedean aspects were already developed in [You17], so we can largely quote those results. For brevity, we sketch the proof in Section 13.

The analogous results on the holomorphic forms (Theorem 1.5) are also similar to the Maass form cases, so we briefly sketch the necessary changes in Section 13.

1.4. *Convention.* The notation $A \ll B$ for quantities A and B means that there exists a constant K such that $|A| \leq KB$ for all relevant A and B, the value of which in each instance should be clear from context. If $p_1, \ldots, p_n$ are parameters, then $\ll_{p_1,\ldots,p_n}$ indicates that the constant K may depend on $p_1, \ldots, p_n$. Implied constants also depend on the choices of implied constants already established in the proof, but we suppress this from this notation. For example, if $|A| \leq K_1 B$ and $|B| \leq K_2 C$, then $A \ll C$ with $K_3 = K_1 K_2$. A major purpose of this notation is to avoid excessive labelling of implied constants. The appearance of the parameter ε among the p_i plays a similar role: each of these ε represents a quantity ε_j that may be taken to be arbitrarily small, and that may depend on all previous $\varepsilon_1, \ldots, \varepsilon_{j-1}$ appearing in the proof.

1.5. *Acknowledgements.* We would like to thank Emmanuel Kowalski for explaining his work on ℓ-adic trace functions to us, which plays a crucial role in Section 9.1 of this paper. We also thank Philippe Michel for pointing out an oversight in an earlier version of that section of the paper and for proposing a solution to it. Part of this work was accomplished during our visit to the Hausdorff Center in Bonn for the summer school on L-functions in 2018. We thank the Center for its support. Lastly, we thank the referees for their careful and thorough reading of this paper.

2. **Automorphic forms and L-functions**

2.1. *Cusp forms.* Let q be a positive integer, and let ψ be a Dirichlet character modulo q. For $t_j \in \mathbb{R} \cup i[-1/2, 1/2]$, let $S_{it_j}(q, \psi)$ be the space of

Maass cusp forms of level q, central character ψ, and spectral parameter t_j. Similarly, for $\kappa \geq 2$, we let $S_\kappa(q, \psi)$ be the space of holomorphic cusp forms of weight κ. Any $f \in S_{it_j}(q, \psi)$ admits a Fourier expansion

$$(2.1) \qquad f(z) = 2\sqrt{y} \sum_{n \neq 0} \lambda_f(n) e(nx) K_{it_j}(2\pi|n|y),$$

and similarly, if $f \in S_\kappa(q, \psi)$, we may write

$$(2.2) \qquad f(z) = \sum_{n=1}^{\infty} \lambda_f(n) n^{\frac{\kappa-1}{2}} e(nz).$$

Now let $\mathcal{H}_{it_j}(m, \psi)$ be the set of Hecke-Maass *newforms* of level $m|q$, normalized so that $\lambda_f(1) = 1$, and define similarly $\mathcal{H}_\kappa(m, \psi)$. Recall the Petersson inner product on $S_{it_j}(q, \psi)$ or $S_\kappa(q, \psi)$ defined by

$$\langle f, g \rangle_q := \int_{\Gamma_0(q)\backslash\mathbb{H}} y^\kappa f(z)\overline{g}(z)\frac{dxdy}{y^2},$$

where in the former case we take $\kappa = 0$. With this normalization of the inner product, we have for any $f \in \mathcal{H}_{it_j}(m, \psi)$ or $\mathcal{H}_\kappa(m, \psi)$ by Rankin-Selberg theory and work of Iwaniec and Hoffstein-Lockhart [Iwa90], [HL94] that

$$(2.3) \quad \langle f, f \rangle_q = \frac{q}{\cosh(\pi t_j)}(q(1 + |t_j|))^{o(1)}, \quad \text{or} \quad \langle f, f \rangle_q = \frac{q\Gamma(\kappa)}{(4\pi)^{\kappa-1}}(q\kappa)^{o(1)}.$$

In fact, we only use the upper bounds implicit in (2.3), which are due to Iwaniec.

Any newform $f \in \mathcal{H}_*(m, \psi)$ satisfies the Hecke relation

$$(2.4) \qquad \lambda_f(n_1)\lambda_f(n_2) = \sum_{d|(n_1,n_2)} \lambda_f(n_1 n_2/d^2)\psi(d).$$

Recall that a Hecke-Maass newform f is called even if $\lambda_f(-1) = 1$, and odd if $\lambda_f(-1) = -1$. It is easy to see that the parity of $f \otimes \chi$ is the parity of f times the parity of χ.

By Atkin-Lehner-Li theory [AL70], [AL78] we have the following direct sum decomposition:

$$(2.5) \qquad S_{it_j}(q, \psi) = \bigoplus_{\ell m = q} \bigoplus_{f \in \mathcal{H}_{it_j}(m,\psi)} S_{it_j}(\ell, f, \psi),$$

where $S_{it_j}(\ell, f, \psi) = \mathrm{span}\{f(dz) : d|\ell\}$, and similarly for holomorphic forms, where each instance of it_j is replaced by κ. The direct sums in (2.5) are orthogonal with respect to the Petersson inner product.

For any $f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)$ with $m|q$, we have by [JL70, Prop. 3.8(iii)] that $f \otimes \chi \in \mathcal{H}_{it_j}(q^2, 1)$, and similarly for holomorphic forms. See also [AL78, Thm. 3.1(ii)] for a classical proof of this fact.

2.2. *Eisenstein series.* Let

(2.6)
$$E_{\chi_1,\chi_2}(z, 1/2 + it) = e_{\chi_1,\chi_2}(y, 1/2 + it) + 2\sqrt{y} \sum_{n \neq 0} \lambda_{\chi_1,\chi_2,t}(n) e(nx) K_{it}(2\pi |n|y),$$

where χ_1, χ_2 are primitive Dirichlet characters modulo q_1, q_2, respectively,

$$\lambda_E(n) = \lambda_{\chi_1,\chi_2,t}(n) = \chi_2(\text{sgn}(n)) \sum_{ab=|n|} \chi_1(a)\overline{\chi_2}(b)a^{-it}b^{it},$$

and $e_{\chi_1,\chi_2}(y, s) = cy^s + c'y^{1-s}$, for certain constants c, c'. Note that the definition (2.6) corresponds to the "completed" Eisenstein series $E^*_{\chi_1,\chi_2}(z, 1/2 + it)$ in [You19], so some care is needed when we quote results from that reference. Then E_{χ_1,χ_2} is of level $m = q_1 q_2$ and central character $\chi_1\overline{\chi_2}$, and is an eigenfunction of all the Hecke operators, and so (2.4) also holds for $\lambda_E(n)$. These are, by definition, the newform Eisenstein series. For two arbitrary Dirichlet characters χ and ψ, let us write $\chi \simeq \psi$ if the underlying primitive characters of χ and ψ are equal. With this notation, we denote the set of newform Eisenstein series by

$$\mathcal{H}_{it,\text{Eis}}(m, \psi) = \{E_{\chi_1,\chi_2}(z, 1/2 + it) : q_1 q_2 = m \text{ and } \chi_1\overline{\chi_2} \simeq \psi\}.$$

In particular, if $E \in \mathcal{H}_{it,\text{Eis}}(m, \psi)$, then $\lambda_E(1) = 1$ and the Hecke relations hold for $\lambda_E(n)$ exactly as they do for $\lambda_f(n)$.

The space $\mathcal{E}_{it}(q, \psi)$, for $t \neq 0$, admits a formal inner product $\langle \cdot, \cdot \rangle_{\text{Eis}}$ induced by

$$\frac{1}{4\pi} \langle E_{\mathfrak{a}}(z, 1/2 + it, \psi), E_{\mathfrak{b}}(z, 1/2 + it, \psi) \rangle_{\text{Eis}} = \delta_{\mathfrak{a}=\mathfrak{b}}.$$

With this definition of the inner product, we have in perfect analogy to (2.3) that

$$(2.7) \quad \langle E_{\chi_1,\chi_2}(z, 1/2 + it), E_{\chi_1,\chi_2}(z, 1/2 + it) \rangle_{\text{Eis}} = \frac{q^{1+o(1)}}{\cosh(\pi t)} |L(1 + 2it, \chi_1\chi_2)|^2.$$

This equation can be deduced from [You19, (8.13), (8.10)], keeping in mind the normalization of the completed Eisenstein series (see [You19, §4]).

There exists an Atkin-Lehner-Li theory for the space $\mathcal{E}_{it}(q, \psi)$, for $t \neq 0$, and a decomposition into spaces of oldforms completely analogous to (2.5). This decomposition is orthogonal with respect to $\langle \cdot, \cdot \rangle_{\text{Eis}}$ and is explained thoroughly in [You19, §8].

Lastly, we define, for $\chi_1\overline{\chi_2} \simeq \overline{\chi}^2$ with χ primitive of conductor q,

$$(2.8) \quad L(s, E_{\chi_1,\chi_2,t} \otimes \chi) = \sum_{n=1}^{\infty} \frac{\lambda_{\chi_1,\chi_2,t}(n)\chi(n)}{n^s} = L(s + it, \chi\chi_1)L(s - it, \chi\overline{\chi_2}).$$

We claim that (2.8) defines the true automorphic L-function of conductor q^2. To see this, check that locally all the solutions to $\chi_1\overline{\chi_2} \simeq \overline{\chi}^2$ with $q_1 q_2 | q$ arise

from $\chi_1 = 1, \chi_2 = \chi^2$ or $\chi_2 = 1$, $\chi_1 = \overline{\chi}^2$. Hence both $\chi\chi_1$ and $\chi\overline{\chi_2}$ are primitive of conductor q.

2.3. *Root Numbers.* Although the theorems in this paper do not depend on the precise values of the root numbers of the forms $f \otimes \chi$, formulas for these are useful when interpreting the main results of this paper. If χ is primitive modulo q, $m \mid q$ and $f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)$ or $f \in \mathcal{H}_{it,\mathrm{Eis}}(m, \overline{\chi}^2)$, then the root number $\epsilon(f \otimes \chi)$ is equal to the parity of f. If $f \in \mathcal{H}_\kappa(m, \overline{\chi}^2)$, then $\epsilon(f \otimes \chi) = i^{-\kappa}\chi(-1)$. These formulas follow from local computations at finite primes using [JL70, Prop. 3.8(iii)] and the explicit formulas for root numbers at the archimedean place found just above [JL70, Thm. 5.15]. See also [Li80, §1].

2.4. *Bruggeman-Kuznetsov.* Let $B_{it_j}(q, \psi)$ denote an orthogonal basis for $S_{it_j}(q, \psi)$, and let $B_{it,\mathrm{Eis}}(q, \psi)$ denote an orthogonal basis for $\mathcal{E}_{it}(q, \psi)$ when $t \neq 0$. Let $h(t)$ be a function holomorphic in the strip $|\mathrm{Im}(t)| \leq \frac{1}{2} + \delta$, satisfying $h(t) = h(-t)$, and $|h(t)| \ll (1 + |t|)^{-2-\delta}$ for some $\delta > 0$. Recall the twisted Kloosterman sum

$$S_\psi(m, n; c) = \sum_{y \,(\mathrm{mod}\ c)}^{*} \overline{\psi}(y) e\Big(\frac{my + n\overline{y}}{c}\Big),$$

where the $*$ on the sum indicates that $(y, c) = 1$, and let $c_t = \frac{4\pi}{\cosh(\pi t)}$. Then, for $mn > 0$, we have (see, e.g., [You19, (10.2)])

$$\sum_{t_j} h(t_j) c_{t_j} \sum_{f \in B_{it_j}(q,\psi)} \frac{\lambda_f(m)\overline{\lambda_f(n)}}{\langle f, f \rangle_q}$$

$$+ \frac{1}{4\pi} \int_{-\infty}^{\infty} h(t) c_t \sum_{E \in B_{it,\mathrm{Eis}}(q,\psi)} \frac{\lambda_E(m)\overline{\lambda_E(n)}}{\langle E, E \rangle_{\mathrm{Eis}}} \, dt$$

$$= \delta_{m=n} g_0 + \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \frac{S_\psi(m, n; c)}{c} g^+\Big(\frac{4\pi\sqrt{mn}}{c}\Big),$$

where

(2.9)
$$g_0 = \frac{1}{\pi} \int_{-\infty}^{\infty} t \tanh(\pi t) h(t) \, dt \quad \text{and} \quad g^+(x) = 2i \int_{-\infty}^{\infty} \frac{J_{2it}(x)}{\cosh(\pi t)} th(t) \, dt.$$

It was shown by the first author [Pet18, §7] that there exist certain positive weights $\rho_f(\ell) = \ell^{o(1)}$ such that if $(n_1 n_2, q) = 1$, then

$$(2.10) \qquad \sum_{\ell m = q} \sum_{f \in \mathcal{H}_{it_j}(m,\psi)} \frac{\lambda_f(n_1)\overline{\lambda_f(n_2)}}{\langle f, f \rangle_q} \frac{1}{\rho_f(\ell)} = \sum_{f \in B_{it_j}(q,\psi)} \frac{\lambda_f(n_1)\overline{\lambda_f(n_2)}}{\langle f, f \rangle_q}.$$

The weight $\rho_f(\ell)$ is a certain explicit function of the Hecke eigenvalues of f. It was shown by the second author [You19, §8.5 and Lem. 8.3] that an analogous

formula holds for the Eisenstein series, namely,

$$(2.11) \quad \sum_{\ell m = q} \sum_{E \in \mathcal{H}_{it,\mathrm{Eis}}(m,\psi)} \frac{\lambda_E(n_1)\overline{\lambda_E(n_2)}}{\langle E, E\rangle_{\mathrm{Eis}}} \frac{1}{\rho_E(\ell)} = \sum_{E \in B_{it,\mathrm{Eis}}(q,\psi)} \frac{\lambda_E(n_1)\overline{\lambda_E(n_2)}}{\langle E, E\rangle_{\mathrm{Eis}}},$$

where $\rho_E(\ell)$ is given by the same function of the Hecke eigenvalues of E as $\rho_f(\ell)$.

Let

$$w_{f,\ell} = c_{t_j} \frac{1}{\langle f, f\rangle_q} \frac{1}{\rho_f(\ell)} \quad \text{and} \quad w_{E,\ell} = c_t \frac{1}{\langle E, E\rangle_{\mathrm{Eis}}} \frac{1}{\rho_E(\ell)}$$

for $f \in \mathcal{H}_{it_j}(m,\psi)$ and $E \in \mathcal{H}_{it,\mathrm{Eis}}(m,\psi)$. Note that $c_{t_j} > 0$ for any $f \in \mathcal{H}_{it_j}(m,\psi)$, including any exceptional cases where $t_j \in i\mathbb{R}$. More precisely, we have by (2.3) (see also [You19, §§8.4, 8.5] for the Eisenstein case)

$$(2.12) \quad w_{f,\ell} = q^{-1}(q(1+|t_j|))^{o(1)} \quad \text{and} \quad w_{E,\ell} = \frac{q^{-1}(q(1+|t|))^{o(1)}}{|L(1+2it,\chi_1\chi_2)|^2}.$$

Note that if $\chi_1\chi_2$ is the trivial character, then this weight vanishes to order 2 at $t = 0$, which is the situation encountered in [CI00]. Indeed, there q is square-free and χ is quadratic, hence the only solution to $\chi_1\overline{\chi_2} \simeq \overline{\chi}^2$ with $q_1 q_2 | q$ is $q_1 = q_2 = 1$, $\chi_1 = \chi_2 = 1$. By the hypothesis in Theorem 1.1 that χ is not quadratic, we have $\chi_1\chi_2$ is not trivial (see the discussion following (2.8)), and hence $w_{E,\ell} \gg q^{-1}(q(1+|t|))^{-\varepsilon}$ for all $t \in \mathbb{R}$. This is the only place where the hypothesis that χ is not quadratic is used in this paper, which is for convenience of notation only.

In summary, we have established the following.

PROPOSITION 2.1. *Suppose χ is primitive of conductor q, and not quadratic. There exist positive weights $w_{f,\ell} \gg q^{-1}(q(1+|t_j|))^{-\varepsilon}$, and $w_{E,\ell} \gg q^{-1}(q(1+|t|))^{-\varepsilon}$ so that for any $(n_1 n_2, q) = 1$ and $n_1 n_2 > 0$, we have*

$$\sum_{t_j} h(t_j) \sum_{\ell m = q} \sum_{f \in \mathcal{H}_{it_j}(m,\overline{\chi}^2)} w_{f,\ell}\lambda_f(n_1)\overline{\lambda_f(n_2)}$$

$$(2.13) \qquad + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(t) \sum_{\ell m = q} \sum_{E \in \mathcal{H}_{it,\mathrm{Eis}}(m,\overline{\chi}^2)} w_{E,\ell}\lambda_E(n_1)\overline{\lambda_E(n_2)}dt$$

$$= \delta_{n_1 = n_2}g_0 + \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \frac{S_{\overline{\chi}^2}(n_1, n_2; c)}{c} g^+\left(\frac{4\pi\sqrt{n_1 n_2}}{c}\right).$$

We also need the opposite-sign case of Proposition 2.1, i.e., when $n_1 n_2 < 0$. The formula is identical to (2.13) except that $g^+(x)$ is replaced by $g^-(x)$ defined by

$$(2.14) \qquad g^-(x) = \frac{4}{\pi} \int_0^{\infty} K_{2it}(x)\sinh(\pi t)th(t)\,dt.$$

3. **Conventions and terminology for weight functions**

We begin with a useful definition from [KPY19]. Let $\mathcal{F}$ be an index set and $X = X_T : \mathcal{F} \to \mathbb{R}_{\geq 1}$ be a function of $T \in \mathcal{F}$.

Definition 3.1. A family $\{w_T\}_{T \in \mathcal{F}}$ of smooth functions supported on a product of dyadic intervals in $\mathbb{R}_{>0}^d$ is called X-*inert* if for each $\mathbf{a} \in \mathbb{Z}_{\geq 0}^d$, we have

$$C(\mathbf{a}) := \sup_{T \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}_{>0}^d} X_T^{-\mathbf{a} \cdot \mathbf{1}} \left| \mathbf{t}^{\mathbf{a}} w_T^{(\mathbf{a})}(\mathbf{t}) \right| < \infty.$$

It is also convenient for later purposes to slightly generalize the above notion of a family of X-inert functions.

Definition 3.2. Suppose that $W_T(x, \mathbf{t})$ with $T \in \mathcal{F}$ is a family of smooth functions, where $\mathbf{t} \in \mathbb{R}^d$. We say that $\{W_T\}_{T \in \mathcal{F}}$ forms an X-inert family with respect to $\mathbf{t}$ if W has dyadic support in terms of $\mathbf{t}$ and if for each $\mathbf{a}$, k and x, we have

$$C_k(x, \mathbf{a}) := \sup_{T \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}_{>0}^d} X_T^{-\mathbf{a} \cdot \mathbf{1}} \left| \mathbf{t}^{\mathbf{a}} \frac{\partial^{\mathbf{a}}}{\partial \mathbf{t}^{\mathbf{a}}} \frac{\partial^k}{\partial x^k} W_T(x, \mathbf{t}) \right| < \infty.$$

As a convention, we may write $w(x, \cdot)$ as shorthand to represent $w(x, \mathbf{t})$. We may then state that $w(x, \cdot)$ is X-inert with respect to $\mathbf{t}$, which allows us to concisely track the behavior of w with respect to the suppressed variables.

4. **Setting up the moment problem**

For $T \geq 1$, let

$$(4.1) \qquad\qquad h_0(t) = \exp(-(t/T)^2) \frac{(t^2 + \frac{1}{4})}{T^2}.$$

Note $h_0(t) > 0$ for $t \in \mathbb{R}$ as well as $-\frac{1}{2} < it < \frac{1}{2}$. Moreover, $h_0(t) \gg T^{-2}$ for $t \ll T$.

In this paper we are concerned with estimating the following moment of L-functions:

$$
\begin{aligned}
\mathcal{M}(q, \chi) := &\sum_{t_j} h_0(t_j) \sum_{\ell m = q} {\sum_{f \in \mathcal{H}_{it_j}(m, \overline{\chi}^2)}}^{+} w_{f, \ell} L(1/2, f \otimes \chi)^3 \\
&+ \frac{1}{4\pi} \int_{-\infty}^{\infty} h_0(t) \sum_{\ell m = q} {\sum_{E \in \mathcal{H}_{it, \mathrm{Eis}}(m, \overline{\chi}^2)}}^{+} w_{E, L} L(1/2, E \otimes \chi)^3 dt,
\end{aligned}
$$

(4.2)

where the $+$ over the sums represents Maass forms or Eisenstein series with even parity.

THEOREM 4.1. *If χ has cube-free conductor and is not quadratic, then we have*

$$\mathcal{M}(q,\chi) \ll_\varepsilon T^B q^\varepsilon.$$

Theorem 4.1 implies Theorem 1.1. Indeed, let $\chi_1 = 1$ and χ_2 be the primitive character underlying χ^2. Then $E = E_{\chi_1,\chi_2,t}$ occurs in $\mathcal{H}_{it,\mathrm{Eis}}(m,\overline{\chi}^2)$ for some $m \mid q$, and we have for this E that

$$L(1/2, E \otimes \chi) = |L(1/2 + it, \chi)|^2.$$

We have as well that $L(1/2, f \otimes \chi) \geq 0$ by [Guo96] (see also (2.8) for the non-negativity in the Eisenstein case), so that Theorem 1.1 follows from Theorem 4.1 by (2.12).

4.1. *Approximate functional equation.* For $j = 1, 2$, let

(4.3)
$$V_j(y,t) = \frac{1}{2\pi i} \int_{(\sigma)} y^{-s} \frac{\Gamma_\mathbb{R}(1/2 + \delta + s + it)^j \Gamma_\mathbb{R}(1/2 + \delta + s - it)^j}{\Gamma_\mathbb{R}(1/2 + \delta + it)^j \Gamma_\mathbb{R}(1/2 + \delta - it)^j} \frac{G_j(s)}{s} ds,$$

where $\Gamma_\mathbb{R}(s) = \pi^{-s/2}\Gamma(s/2)$, $\delta \in \{0,1\}$, and σ is to the right of all poles of the integrand. We take $G_1(s) = e^{2s^2}$ and $G_2(s) = e^{4s^2}$. Here $V_j(x,t)$ is a smooth function on $x > 0$ with rapid decay for $x \gg 1 + |t|^j$. See Section 10 for more precise estimates for V_j.

LEMMA 4.2. *Suppose $m \mid q$ and $f \in \mathcal{H}_{it_j}(m,\overline{\chi}^2)$ is even. We have*

$$L(1/2, f \otimes \chi)^3$$
$$= \sum_{(d,q)=1} \frac{4}{d} \sum_{n_1,n_2,n_3} \frac{\lambda_f(n_1)\overline{\lambda_f}(n_2 n_3)\chi(n_1)\overline{\chi}(n_2 n_3)}{\sqrt{n_1 n_2 n_3}} V_1\left(\frac{n_1}{q}, t_j\right) V_2\left(\frac{n_2 n_3 d^2}{q^2}, t_j\right),$$

and similarly for $L(1/2, E \otimes \chi)^3$ for $m \mid q$ and $E \in \mathcal{H}_{it,\mathrm{Eis}}(m,\overline{\chi}^2)$ even. The parity parameter δ implicit in the definition of V_j is equal to the parity of χ.

Proof. Since f is even, the root number $\epsilon(f \otimes \chi)$ is $+1$. For f a Maass newform of spectral parameter t_j, a standard approximate functional equation [IK04, Thm. 5.3] gives

(4.4)
$$L(1/2, f \otimes \chi) = 2\sum_{n_1} \frac{\lambda_f(n_1)\chi(n_1)}{\sqrt{n_1}} V_1\left(\frac{n_1}{q}, t_j\right),$$

where $\delta = 0$ if χ is even and $\delta = 1$ if χ is odd. Similarly we have

(4.5)
$$L(1/2, f \otimes \chi)^2 = 2 \sum_{(d,q)=1} \frac{1}{d} \sum_{n_2,n_3} \frac{\overline{\lambda_f}(n_2 n_3)\overline{\chi}(n_2 n_3)}{\sqrt{n_2 n_3}} V_2\left(\frac{n_2 n_3 d^2}{q^2}, t_j\right),$$

where the conjugates appear for convenience since $\lambda_f(n)\chi(n) \in \mathbb{R}$, and the sum over d arises from the Hecke relation (2.4).

The product of (4.4) and (4.5) gives the formula in the statement of the lemma. □

4.2. *Bruggeman-Kuznetsov.* Let $w_0(\cdot) = w_0(n_1, n_2, n_3, c)$ be a family of 1-inert functions (depending on q, T, N_j, C) with dyadic support on $n_j \asymp N_j$ and $c \asymp C$, and let $N_1, N_2, N_3, C \gg 1$. Define $J_0^\pm = J_0^\pm(x, n_1, n_2, n_3, c)$ by

$$(4.6) \qquad J_0^+(x, n_1, n_2, n_3, c) = w_0(\cdot) \int_{-\infty}^{\infty} \frac{J_{2it}(x)}{\cosh(\pi t)} th\left(t, \frac{n_1}{q}, \frac{n_2 n_3 d^2}{q^2}\right) dt,$$

with J_0^- defined similarly with $\frac{J_{2it}(x)}{\cosh(\pi t)}$ replaced by $K_{2it}(x)\sinh(\pi t)$, where in both cases

$$(4.7) \qquad h(t, y_1, y_2) = \exp(-(t/T)^2)\frac{(t^2 + \frac{1}{4})}{T^2}V_1(y_1, t)V_2(y_2, t).$$

Let $\mathcal{S}_{N_1, N_2, N_3, C}^\pm$ be defined by

$$\frac{1}{C\sqrt{N_1 N_2 N_3}} \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \sum_{n_1, n_2, n_3} \chi(n_1)\overline{\chi}(n_2 n_3)S_{\overline{\chi}^2}(n_1, n_2 n_3; c)J_0^\pm\left(\frac{4\pi\sqrt{n_1 n_2 n_3}}{c}, \cdot\right).$$

PROPOSITION 4.3. *Suppose that there exists $B > 2$ such that $\mathcal{S}_{N_1, N_2, N_3, C}^\pm$ $\ll_\varepsilon T^B q^\varepsilon$ for all N_1, N_2, N_3, C satisfying*

$$(4.8) \qquad N_1 \ll_\varepsilon (qT)^{1+\varepsilon}, \qquad N_2 N_3 \ll_\varepsilon d^{-2}(qT)^{2+\varepsilon}, \qquad q \ll C \ll (qT)^{100}.$$

Then Theorem 4.1 holds.

Proof. Recall the even parity condition on the sums over newforms in (4.2). This condition can be detected by extending the sums to all newforms and inserting the indicator function $\frac{1}{2}(1 + \lambda_f(-1))$ for Maass forms and Eisenstein series. By (2.13), we have

$$\mathcal{M}(q, \chi) = \mathcal{D} + \tfrac{1}{2}\mathcal{S}^+ + \tfrac{1}{2}\mathcal{S}^-,$$

where $\mathcal{D}$ is the diagonal term, and

$$
\begin{aligned}
(4.9) \qquad \mathcal{S}^\pm = &\sum_{(d,q)=1} \frac{4}{d} \sum_{n_1, n_2, n_3} \frac{\chi(n_1)\overline{\chi}(n_2 n_3)}{\sqrt{n_1 n_2 n_3}} \\
&\times \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \frac{S_{\overline{\chi}^2}(\pm n_1, n_2 n_3; c)}{c} g^\pm\left(\frac{4\pi\sqrt{n_1 n_2 n_3}}{c}\right).
\end{aligned}
$$

Here $g^\pm(x)$ is defined by (2.9) and (2.14) with respect to $h(t, \frac{n_1}{q}, \frac{n_2 n_3 d^2}{q^2})$ defined in (4.7).

The function h is a valid test function for the hypotheses in the Bruggeman-Kuznetsov formula, and one may derive a crude bound of the form $g^\pm(x) \ll x^{1-\varepsilon}T^{1+\varepsilon}$, as we will show in Section 10. Hence by the Weil bound (see, e.g., [KL13, Thm. 9.2], which gives $|S_\psi(a, b; c)| \le d(c)(a, b, c)^{1/2}c^{1/2}q^{1/2}$, where

ψ has conductor $q|c$ and $d(\cdot)$ is the divisor function), we have that the sum over c in (4.9) converges absolutely. We further develop the analytic properties of $g^{\pm}(x)$ in Section 10.

It is easy to see that $\mathcal{D} \ll_{\varepsilon} T^{2+\varepsilon}q^{\varepsilon}$, and so the proof of Proposition 4.3 reduces to showing that $\mathcal{S}^{\pm} \ll_{\varepsilon} T^{B}q^{\varepsilon}$.

Next we apply a dyadic partition of unity to each of n_1, n_2, n_3, c. Consider the component $w_0(\cdot)$ of this partition of unity that localizes the variables by $n_j \asymp N_j$, $c \asymp C$. We may assume the inequalities (4.8) hold, since if they do not, then the contribution from that piece of the partition of unity is small by trivial bounds. Hence,

$$(4.10) \qquad \mathcal{S}^{\pm} = \sum_{(d,q)=1} \frac{4}{d} \sum_{N_1,N_2,N_3,C} \mathcal{S}^{\pm}_{N_1,N_2,N_3,C} + O_{\varepsilon}((qT)^{\varepsilon}),$$

where N_1, N_2, N_3, C run over dyadic number satisfying the bounds (4.8). From the hypothesis on $\mathcal{S}^{\pm}_{N_1,N_2,N_3,C}$ in the statement of the proposition, we conclude the proof. $\qquad \square$

4.3. *Poisson summation.* Let $m_1, m_2, m_3 \in \mathbb{Z}$ and $c > 0$. Let $G = G(m_1, m_2, m_3; c)$ be the character sum defined by

$$G = c^{-3} \sum_{y \,(\mathrm{mod}\ c)}^{*} \sum_{x_1,x_2,x_3 \,(\mathrm{mod}\ c)}$$
$$\times \chi(x_1)\overline{\chi}(x_2 x_3)\chi^2(y)e_c(m_1 x_1 + m_2 x_2 + m_3 x_3 + x_1 y + x_2 x_3 \overline{y}),$$

where $e_c(x) = e(x/c)$.

Let $M_1, M_2, M_3 > 0$, and let $w(\cdot) = w(n_1, n_2, n_3, c, m_1, m_2, m_3)$ be a family of 1-inert functions (depending on q, T, N_j, C, M_j) with dyadic support on $n_j \asymp N_j$, $c \asymp C$, and $m_j \asymp M_j$. Let $J^{\pm}(x, \cdot) = J^{\pm}(x, n_1, n_2, n_3, m_1, m_2, m_3, c)$ be defined by
(4.11)
$$J^{+}(x, n_1, n_2, n_3, m_1, m_2, m_3, c) = w(\cdot) \int_{-\infty}^{\infty} \frac{J_{2it}(x)}{\cosh(\pi t)} th\left(t, \frac{n_1}{q}, \frac{n_2 n_3 d^2}{q^2}\right) dt,$$

and let J^{-} be defined similarly with $K_{2it}(x)\sinh(\pi t)$ in place of $\frac{J_{2it}(x)}{\cosh(\pi t)}$. Note that $J^{\pm}$ is identical to $J_0^{\pm}$ except that $w_0(\cdot)$ is replaced by $w(\cdot)$, which depends on the additional variables m_1, m_2, m_3.

Let
(4.12)
$$K_0^{\pm} = \int_{\mathbb{R}^3} J_0^{\pm}\left(\frac{4\pi\sqrt{t_1 t_2 t_3}}{c}, t_1, t_2, t_3, c\right) e_c(-m_1 t_1 - m_2 t_2 - m_3 t_3) dt_1 dt_2 dt_3$$

and
(4.13)
$$K^{\pm} = \int_{\mathbb{R}^3} J^{\pm}\Big(\frac{4\pi\sqrt{t_1 t_2 t_3}}{c}, t_1, t_2, t_3, \cdot\Big) e_c(-m_1 t_1 - m_2 t_2 - m_3 t_3) dt_1 dt_2 dt_3.$$

Finally, let $\epsilon_1, \epsilon_2, \epsilon_3 \in \{\pm 1\}$,

$$(4.14) \quad \mathcal{T}^{\pm} = \mathcal{T}^{\pm}_{\epsilon_1, \epsilon_2, \epsilon_3} = \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \sum_{m_j \epsilon_j \geq 1} G(m_1, m_2, m_3; c) \frac{K^{\pm}(m_1, m_2, m_3, c)}{C\sqrt{N_1 N_2 N_3}},$$

and

$$(4.15) \qquad \mathcal{T}^{\pm}_0 = \sum_{c \equiv 0 \,(\mathrm{mod}\ q)} \sum_{m_1 m_2 m_3 = 0} G(m_1, m_2, m_3; c) \frac{K^{\pm}_0(m_1, m_2, m_3, c)}{C\sqrt{N_1 N_2 N_3}}.$$

PROPOSITION 4.4. *Suppose that* $\mathcal{T}^{\pm}, \mathcal{T}^{\pm}_0 \ll_{\varepsilon} T^B q^{\varepsilon}$ *for some* $B > 2$ *and for all* N_1, N_2, N_3, C *satisfying* (4.8) *and all* M_1, M_2, M_3 *satisfying* $M_j \ll (qT)^A$ *for some large but fixed* A. *Then* $\mathcal{S}^{\pm}_{N_1, N_2, N_3, C} \ll_{\varepsilon} T^B q^{\varepsilon}$ *for all such* N_1, N_2, N_3, C.

Sections 5–12 are dedicated to the proof of the bounds $\mathcal{T}^{\pm}, \mathcal{T}^{\pm}_0 \ll_{\varepsilon} T^B q^{\varepsilon}$, which by Propositions 4.4 and 4.3 will finish the proof of Theorem 4.1, and hence of Theorem 1.1.

Proof. Applying Poisson summation in each of the variables n_1, n_2, n_3 modulo c gives

(4.16)
$$\mathcal{S}^{\pm}_{N_1, N_2, N_3, C} = \frac{1}{C\sqrt{N_1 N_2 N_3}} \sum_{c \equiv 0 \,(\mathrm{mod}\ q)}$$
$$\times \sum_{m_1, m_2, m_3 \in \mathbb{Z}} \chi(\pm 1) G(\pm m_1, m_2, m_3; c) K^{\pm}_0(m_1, m_2, m_3, c).$$

By integrating $K^{\pm}_0$ by parts three times in each variable, we have by (4.8) a crude bound of the form

$$(4.17) \qquad K^{\pm}_0(m_1, m_2, m_3, c) \ll (qT)^A \prod_{j=1}^{3} (1 + |m_j|)^{-3}$$

for some possibly large but fixed A. Therefore the sum (4.16) converges absolutely, and we may in fact truncate each m_j variable at $|m_j| \ll (qT)^{A'}$ for some large A' depending polynomially on $1/\varepsilon$ at the cost of a small error term.

Next, we separate the terms with $m_1 m_2 m_3 = 0$ in $\mathcal{S}^{\pm}_{N_1, N_2, N_3, C}$ from those in which none of the m_j vanish. The terms with $m_1 m_2 m_3 = 0$ form the sum $\mathcal{T}^{\pm}_0$ defined in (4.15). Leaving these terms aside, we split the remaining terms for which $m_j \neq 0$ for all j into eight separate sums according to the

octants of $\mathbb{Z}^3 - \{m_1 m_2 m_3 = 0\}$. Let us parametrize these eight sums by $(\epsilon_1, \epsilon_2, \epsilon_3) \in \{\pm 1\}^3$. The octant corresponding to $\epsilon_1, \epsilon_2, \epsilon_3$ is then described by the inequalities $m_j \epsilon_j \geq 1$ for $j = 1, 2, 3$. Given one choice of signs $\epsilon_j \in \{\pm 1\}$, we insert a dyadic partition of unity to the m_1, m_2, m_3 sums, which localizes each $|m_j| \asymp M_j \ll (qT)^{A'}$. The result of all of these decompositions is that

$$\mathcal{S}^{\pm}_{N_1, N_2, N_3, C} = \mathcal{T}_0^{\pm} + \sum_{M_1, M_2, M_3} \sum_{\epsilon_1, \epsilon_2, \epsilon_3 \in \{\pm 1\}} \mathcal{T}^{\pm}_{\epsilon_1, \epsilon_2, \epsilon_3} + O_\varepsilon((qT)^\varepsilon).$$

The proposition now follows from the hypothesized bounds on $\mathcal{T}_0^{\pm}$ and $\mathcal{T}^{\pm}$.   □

The main focus in this paper is on the character sum G, which is a generalization of the character sum found in the previous works [CI00], [You17], [Pet15], [PY19b], since χ is no longer assumed to be quadratic and q is not necessarily square-free. On the other hand, $K^{\pm}$ is very similar in shape to the oscillatory integrals found in the above references, so in Section 11 we largely quote the existing literature.

5. The calculation of G

Based on the structural approach presented in [PY19b], our primary goal on the arithmetical aspects of G is to understand the analytic properties of the Dirichlet series

(5.1)
$$Z(s_1, s_2, s_3, s_4)$$
$$:= \sum_{\epsilon_1 m_1, \epsilon_2 m_2, \epsilon_3 m_3 \geq 1} \sum_{c \equiv 0 \,(\mathrm{mod}\, q)} \frac{cq G(m_1, m_2, m_3; c) e_c(-m_1 m_2 m_3) \chi(-1)}{m_1^{s_1} m_2^{s_2} m_3^{s_3} (c/q)^{s_4}}.$$

For simplicity of notation, we only consider the case of (5.1) where $\epsilon_j = 1$ for all j, since the other sign combinations can be treated in the same way. Of course, we cannot neglect to study the contribution from $m_1 m_2 m_3 = 0$ as well. In any event, we calculate G in explicit form as much as possible.

5.1. *Simplifications.* Write $c = qr$ with $r \geq 1$. We have

$$\sum_{x_1 \,(\mathrm{mod}\, c)} \chi(x_1) e_c(m_1 x_1 + x_1 y) = r \tau(\chi) \overline{\chi}\left(\frac{m_1 + y}{r}\right),$$

where the sum vanishes unless $y \equiv -m_1 \pmod{r}$, and $\tau(\chi)$ denotes the Gauss sum of $\chi \pmod{q}$. Similarly, we calculate the x_2 sum by

$$\sum_{x_2 \,(\mathrm{mod}\, c)} \overline{\chi}(x_2) e_c(x_2(m_2 + x_3 \overline{y})) = r \tau(\overline{\chi}) \chi\left(\frac{m_2 + x_3 \overline{y}}{r}\right),$$

where the sum vanishes unless $x_3 \equiv -m_2 y \pmod{r}$. Changing variables $x_3 \to yx_3$, we hence obtain

$$
G = \frac{r^2 \tau(\chi)\tau(\overline{\chi})}{c^3} \sum_{\substack{y \pmod{c} \\ y \equiv -m_1 \pmod{r}}}^{*}
$$

$$
\times \sum_{\substack{x_3 \pmod{c} \\ x_3 \equiv -m_2 \pmod{r}}} \overline{\chi}(x_3)\chi(y)e_c(m_3 y x_3)\overline{\chi}\Big(\frac{m_1 + y}{r}\Big)\chi\Big(\frac{m_2 + x_3}{r}\Big).
$$

Since $(y, c) = 1$, we learn that $G = 0$ unless

$$(5.2) \hspace{4cm} (m_1, r) = 1.$$

Provided we maintain this condition, we can drop the condition that $(y, c) = 1$. Writing $y = -m_1 + ru$ and $x_3 = -m_2 + rt$, we obtain

$$
G(m_1, m_2, m_3; c) = c^{-3} r^2 \tau(\chi)\tau(\overline{\chi}) e_c(m_1 m_2 m_3) H_\chi(m_1, m_2, m_3, r) \delta_{(m_1, r)=1},
$$

where

$$
H_\chi(m_1, m_2, m_3, r) = \sum_{u, t \pmod{q}} \chi(t)\overline{\chi}(u)\overline{\chi}(-m_2 + rt)\chi(-m_1 + ru)
$$

$$
\times e_c(m_3(-m_1 + ru)(-m_2 + rt) - m_1 m_2 m_3).
$$

Note that

$$(5.3) \quad \delta_{(m_1, r)=1} H_\chi(m_1, m_2, m_3, r) = cqG(m_1, m_2, m_3; c)e_c(-m_1 m_2 m_3)\chi(-1),$$

so that

$$(5.4) \hspace{2cm} Z(s_1, s_2, s_3, s_4) = \sum_{\substack{m_1, m_2, m_3, r \geq 1 \\ (m_1, r)=1}} \frac{H_\chi(m_1, m_2, m_3, r)}{m_1^{s_1} m_2^{s_2} m_3^{s_3} r^{s_4}}.$$

Next we derive some simple but useful symmetries of H_χ. Although we only need to compute $H_\chi(m_1, m_2, m_3, r)$ when (5.2) holds, it will be more convenient not to assume this condition. Changing variables $t \to (-m_1 + ru)^{-1} t$ gives

$$
H_\chi(m_1, m_2, m_3, r)
$$

$$
= \sum_{u, t \pmod{q}} \chi(t)\chi(-m_1 + ru)\overline{\chi}(u)\overline{\chi}(rt - m_2(-m_1 + ru))e_q(m_3 t - m_2 m_3 u).
$$

Next shift by $t \to t + m_2 u$, giving
$$(5.5)$$
$$
H_\chi(m_1, m_2, m_3, r) = \sum_{u, t \pmod{q}} \chi(t + m_2 u)\overline{\chi}(rt + m_1 m_2)\overline{\chi}(u)\chi(-m_1 + ru)e_q(m_3 t).
$$

Since $G(m_1, m_2, m_3; c)$ is symmetric in m_2, m_3, we see that

$$(5.6) \qquad H_\chi(m_1, m_2, m_3, r) = H_\chi(m_1, m_3, m_2, r).$$

Observe that if $(m_1 m_2, r, q) \neq 1$, then every summand in (5.5) vanishes. Together with the symmetry (5.6), we obtain

$$(5.7) \qquad H_\chi(m_1, m_2, m_3, r) = 0 \quad \text{if } (m_1 m_2 m_3, r, q) \neq 1.$$

If $(q, r) = 1$, then there is some additional symmetry. We claim that

$$(5.8) \qquad H_\chi(m_1, m_2, m_3, r) = H_{\overline{\chi}}(m_2, m_1, m_3, r) \quad \text{if } (q, r) = 1.$$

Indeed, changing variables $t \to \overline{r}t$, $u \to \overline{r}u$, gives

$$H_\chi(m_1, m_2, m_3, r) = \sum_{u,t \,(\mathrm{mod}\ q)} \chi(t + m_2 u)\overline{\chi}(t + m_1 m_2)\overline{\chi}(u)\chi(-m_1 + u)e_q(m_3 \overline{r}t).$$

Next we change variables $u \to u + m_1$, followed by $t \to ut - m_1 m_2$ (note u is coprime to q for every non-zero summand), giving

$$
\begin{aligned}
&H_\chi(m_1, m_2, m_3, r) \\
(5.9) \quad &= e_q(-m_1 m_2 m_3 \overline{r}) \sum_{u,t \,(\mathrm{mod}\ q)} \chi(t + m_2)\overline{\chi}(t)\overline{\chi}(u + m_1)\chi(u)e_q(m_3 \overline{r}ut),
\end{aligned}
$$

from which we deduce (5.8).

5.2. *Decomposition into Dirichlet characters.* It is possible to calculate H_χ further, as in [CI00], but going to the Fourier transform of H_χ turns out to be a more advantageous move.

Begin by writing $r = r_0 r'$ and $m_j = m_{j,0} m_j'$, $j = 1, 2, 3$, with

$$(5.10) \qquad m_{j,0} | q^\infty, \qquad r_0 | q^\infty$$

and $(m_1' m_2' m_3' r', q) = 1$. Inside the expression (5.5), change variables $t \to m_1' m_2' \overline{r'}t$ and $u \to \overline{r'}m_1' u$, giving

$$H_\chi(m_1, m_2, m_3, r) = H_\chi(m_{1,0}, m_{2,0}, m_{3,0}w, r_0),$$

where

$$w = m_1' m_2' m_3' \overline{r'}.$$

Note that $(w, q) = 1$ by assumption. We may then view H_χ as a function of w on $(\mathbb{Z}/q\mathbb{Z})^\times$ and apply multiplicative Fourier analysis. That is, we write

$$(5.11) \qquad H_\chi(m_{1,0}, m_{2,0}, m_{3,0}w, r_0) = \frac{1}{\varphi(q)} \sum_{\psi \,(\mathrm{mod}\ q)} \widehat{H}(\psi)\psi(w),$$

where

$$
\begin{aligned}
\widehat{H}(\psi) = \widehat{H} &= \widehat{H}(\psi, \chi, m_{1,0}, m_{2,0}, m_{3,0}, r_0) \\
(5.12) \qquad &= \sum_{v \,(\mathrm{mod}\ q)} H_\chi(m_{1,0}, m_{2,0}, m_{3,0}v, r_0)\overline{\psi}(v).
\end{aligned}
$$

Expanding the definition, we have

(5.13)
$$\widehat{H}(\psi, \chi, m_1, m_2, m_3, r)$$

$$= \sum_{t,u,v \,(\mathrm{mod}\ q)} \chi(t + m_2 u)\overline{\chi}(rt + m_1 m_2)\overline{\chi}(u)\chi(-m_1 + ru)e_q(m_3 vt)\overline{\psi}(v).$$

The sum $\widehat{H}(\psi)$ inherits from (5.6) and (5.8) the symmetries

(5.14)
$$\widehat{H}(\psi, \chi, m_1, m_3, m_2, r) = \widehat{H}(\psi, \chi, m_1, m_2, m_3, r),$$

$$\widehat{H}(\psi, \chi, m_2, m_1, m_3, r) = \widehat{H}(\psi, \overline{\chi}, m_1, m_2, m_3, r) \qquad \text{if } (q, r) = 1.$$

Similarly, from (5.7) we deduce

(5.15) $$\widehat{H}(\psi, \chi, m_1, m_2, m_3, r) = 0 \qquad \text{if} \qquad (m_1 m_2 m_3, r, q) \neq 1.$$

We immediately see the pleasant factorization

(5.16) $$Z(s_1, s_2, s_3, s_4) = \frac{1}{\varphi(q)} \sum_{\psi \,(\mathrm{mod}\ q)} \frac{L(s_1, \psi)L(s_2, \psi)L(s_3, \psi)L(s_4, \overline{\psi})}{\zeta^{(q)}(s_1 + s_4)} Z_{\mathrm{fin}},$$

where

$$Z_{\mathrm{fin}} = Z_{\mathrm{fin}}(s_1, s_2, s_3, s_4) := \sum_{\substack{m_{1,0}, m_{2,0}, m_{3,0}, r_0 | q^\infty \\ (m_{1,0}, r_0) = 1}} \frac{\widehat{H}(\psi, \chi, m_{1,0}, m_{2,0}, m_{3,0}, r_0)}{m_{1,0}^{s_1} m_{2,0}^{s_2} m_{3,0}^{s_3} r_0^{s_4}}.$$

The factor $\zeta^{(q)}(s_1 + s_4)^{-1}$ arose from Möbius inversion to detect $(m_1', r') = 1$.

Now the task is to understand the analytic properties of Z_{fin}. Suppose $q = q_1 q_2$ with $(q_1, q_2) = 1$, $\chi = \chi_1 \chi_2$ and $\psi = \psi_1 \psi_2$ with χ_j, ψ_j modulo q_j. Similarly, write $a = a_1 a_2$, and so on with b, c, d. By the Chinese remainder theorem, we have that $\widehat{H}(\psi, \chi, a, b, c, d)$ factors as a sum of modulus q_1 times a sum of modulus q_2. The sum modulo q_1 equals

$$\sum_{t_1, u_1, v_1 \,(\mathrm{mod}\ q_1)} \chi_1(t_1 + b_1 b_2 u_1)\overline{\chi_1}(d_1 d_2 t_1 + a_1 a_2 b_1 b_2)\overline{\chi_1}(u_1)$$

$$\times \chi_1(-a_1 a_2 + d_1 d_2 u_1)e_{q_1}(\overline{q_2}c_1 c_2 v_1 t_1)\overline{\psi_1}(v_1).$$

Changing variables $t_1 \to a_2 b_2 \overline{d_2} t_1$, $u_1 \to a_2 \overline{d_2} u_1$, and $v_1 \to \overline{a_2 b_2 c_2} d_2 q_2 v_1$ shows

$$\widehat{H}(\psi, \chi, a, b, c, d) = \epsilon \widehat{H}(\psi_1, \chi_1, a_1, b_1, c_1, d_1)\widehat{H}(\psi_2, \chi_2, a_2, b_2, c_2, d_2),$$

where $\epsilon = \psi_1(a_2 b_2 c_2 \overline{q_2 d_2})\psi_2(a_1 b_1 c_1 \overline{q_1 d_1})$. Pleasantly, $\widehat{H}$ is almost multiplicative in terms of χ, ψ, and the only "twisted" aspect comes from the factor ϵ.

This shows

$$(5.17) \qquad Z_{\text{fin}} = \omega \prod_{p^j \| q} \sum_{\substack{a,b,c,d | p^\infty \\ (a,d)=1}} \frac{\eta(abc)\overline{\eta}(d)}{a^{s_1} b^{s_2} c^{s_3} d^{s_4}} \widehat{H}(\psi_p, \chi_p, a, b, c, d),$$

where η is some Dirichlet character depending on ψ and p, and ω is some complex number of absolute value 1, which depends on ψ. Here χ_p, ψ_p are the p-parts of χ, ψ.

6. Evaluation of $\widehat{H}$

Here we comprehensively evaluate $\widehat{H}$ when $q = p^k$, $k \geq 1$. Recall that $\widehat{H}$ was defined in (5.13). Throughout this section we assume $m_1, m_2, m_3, r | q^\infty$.

6.1. *Elementary lemmas on character sums.* We begin with some character sum evaluations that are used repeatedly in the calculations of $\widehat{H}$.

LEMMA 6.1. *Suppose that χ is primitive modulo q and $d|q$, $d \neq q$. Then*

$$\sum_{\substack{a \,(\mathrm{mod}\ q) \\ a \equiv b \,(\mathrm{mod}\ d)}} \chi(a) = 0.$$

This well-known lemma may be found in [IK04, (3.9)], for instance.

LEMMA 6.2. *Suppose p is prime, $a \in \mathbb{Z}$, and χ has conductor p^k, $k \geq 2$. Then*

$$\sum_{\substack{t \,(\mathrm{mod}\ p^k) \\ t \equiv a \,(\mathrm{mod}\ p)}} \chi(t)\overline{\chi}(t+1) = 0.$$

Proof. If $(a(a + 1), p) \neq 1$, then the sum is empty, so suppose otherwise. Then from $\chi(t)\overline{\chi}(t+1) = \overline{\chi}(1+\overline{t})$, and changing variables $t \to \overline{t}$, the conclusion follows from Lemma 6.1. $\square$

LEMMA 6.3. *Suppose χ is primitive modulo q, and let $a, b, c, d \in \mathbb{Z}$ with $(a, c, q) = 1$. Then*

$$(6.1) \qquad \sum_{t \,(\mathrm{mod}\ q)} \chi(at + b)\overline{\chi}(ct + d) = \chi(a)\overline{\chi}(c)R_q(ad - bc),$$

where $R_q(n) = S(n, 0; q)$ is the Ramanujan sum.

Proof. We first claim the sum vanishes unless $(a, q) = (c, q) = 1$. By symmetry, suppose $(a, q) \neq 1$. Then $\chi(at + b)$ is constant for t ranging over an arithmetic progression modulo $\frac{q}{(a,q)}$. Lemma 6.1 shows that the sum over this arithmetic progression of $\overline{\chi}(ct + d)$ vanishes unless $q | c\frac{q}{(a,q)}$, i.e., $(a, q)|c$, whence $1 = (a, c, q) = (a, q)$, a contradiction. Therefore, (6.1) is derived if $(a, q) \neq 1$ or $(c, q) \neq 1$.

Now suppose $(a, q) = (c, q) = 1$. By converting to additive characters, that is, using

$$(6.2) \qquad \chi(at + b) = \frac{1}{\tau(\overline{\chi})} \sum_{x \, (\mathrm{mod} \, q)} \overline{\chi}(x) e_q(x(at + b)),$$

and likewise for $\overline{\chi}(ct + d)$, the formula (6.1) follows from a routine calculation. □

6.2. *The case ψ primitive modulo q.*

LEMMA 6.4. *Suppose p is a prime and $q = p^k$, $k \geq 1$. Suppose ψ is primitive modulo q. Then $\widehat{H}$ vanishes unless $(m_1 m_2 m_3 r, q) = 1$, in which case*

$$(6.3) \qquad \widehat{H}(\psi, \chi, 1, 1, 1, 1) = \tau(\overline{\psi}) g(\chi, \psi),$$

where $g(\chi, \psi)$ was defined by (1.9), and $\tau(\overline{\psi})$ is the Gauss sum.

Proof. Since ψ is primitive, the sum over v in (5.13) is a Gauss sum, giving

$$\widehat{H}(\psi, \chi, m_1, m_2, m_3, r)$$
$$= \tau(\overline{\psi}) \sum_{t, u \, (\mathrm{mod} \, q)} \chi(t + m_2 u) \overline{\chi}(rt + m_1 m_2) \overline{\chi}(u) \chi(-m_1 + ru) \psi(m_3 t).$$

Hence, $\widehat{H}$ vanishes unless $(m_3, q) = 1$. By the first symmetry in (5.14), this means it vanishes unless $(m_2, q) = 1$, too. We claim that it vanishes unless $(m_1, q) = 1$. If $p | m_1$ and $(p, r) = 1$, then the claim follows from the second symmetry in (5.14), while if $p | (m_1, r)$, then the claim follows from (5.15). Thus we may set $m_1 = m_2 = m_3 = 1$, since we have assumed that $m_1 m_2 m_3 r | q^\infty$.

If $(p, r) = 1$, then $r = 1$, in which case

$$\widehat{H}(\psi, \chi, 1, 1, 1, 1) = \tau(\overline{\psi}) \sum_{t, u \, (\mathrm{mod} \, q)} \chi(t + u) \overline{\chi}(t + 1) \overline{\chi}(u) \chi(u - 1) \psi(t).$$

Changing variables $u \to u + 1$ followed by $t \to ut - 1$, and finally changing the roles of u and t (for cosmetic purposes), we obtain (6.3).

Finally, suppose that $p | r$ and $m_1 m_2 m_3 = 1$. Changing variables $t \to ut$ gives

$$\widehat{H}(\psi, \chi, 1, 1, 1, r) = \tau(\overline{\psi}) \sum_{t, u \, (\mathrm{mod} \, q)} \chi(1 + t) \chi(-1 + ru) \overline{\chi}(1 + rut) \psi(tu).$$

Since $\chi(-1 + ru) \overline{\chi}(1 + rut)$ is periodic in u with period $\frac{p^k}{(r, p^k)} \leq p^{k-1}$, the sum over u vanishes by Lemma 6.1, since ψ has conductor p^k. □

6.3. *The case of ψ trivial.*

LEMMA 6.5. *Suppose $\psi = \chi_0$ is the trivial character, and $q = p^k$, $k \geq 1$. Then*

$$\widehat{H}(\chi_0, \chi, m_1, m_2, m_3, r) = \chi_0(r)R_q(m_1)R_q(m_2)R_q(m_3)$$
$$+ qR_q(r)\chi(-1)\chi_0(m_1m_2m_3).$$

Proof. In this case, $\widehat{H}(\chi_0, \chi, m_1, m_2, m_3, r)$ equals

$$\sum_{t,u \,(\mathrm{mod}\, q)} \chi(t + m_2u)\overline{\chi}(rt + m_1m_2)\overline{\chi}(u)\chi(-m_1 + ru)R_q(m_3t).$$

Write $R_q(m_3t) = R_q(m_3) + (R_q(m_3t) - R_q(m_3))$, and note that if $p \nmid t$, then $R_q(m_3t) - R_q(m_3) = 0$. We accordingly write $\widehat{H} = S_1 + S_2$, where

$$S_1 = R_q(m_3) \sum_{t,u \,(\mathrm{mod}\, q)} \chi(t + m_2u)\overline{\chi}(rt + m_1m_2)\overline{\chi}(u)\chi(-m_1 + ru),$$

and $S_2 = \widehat{H} - S_1$. We will show

$$S_1 = \chi_0(r)R_q(m_1)R_q(m_2)R_q(m_3) \quad \text{and} \quad S_2 = qR_q(r)\chi(-1)\chi_0(m_1m_2m_3).$$

First we evaluate S_1. By Lemma 6.3 and since we may assume $(m_1 - ru, q) = 1$, we have

$$\sum_{t \,(\mathrm{mod}\, q)} \chi(t + m_2u)\overline{\chi}(rt + m_1m_2) = \overline{\chi}(r)R_q(m_2).$$

To finish the evaluation of S_1, we apply Lemma 6.3 to give

$$\sum_{u \,(\mathrm{mod}\, q)} \overline{\chi}(u)\chi(ru - m_1) = \chi(r)R_q(m_1).$$

Now we evaluate S_2. The t-sum is restricted by $p|t$, and so we see that S_2 vanishes unless $(p, m_1m_2) = 1$. By our convention, we may set $m_1 = m_2 = 1$, giving

$$S_2 = \chi_0(m_1m_2)\chi(-1)$$
$$\times \sum_{\substack{t \,(\mathrm{mod}\, q) \\ p|t}} \sum_{u \,(\mathrm{mod}\, q)} \chi(t + u)\overline{\chi}(rt + 1)\overline{\chi}(u)\chi(1 - ru)(R_q(m_3t) - R_q(m_3)).$$

Next we change variables $t \to ut$, giving

$$S_2 = \chi_0(m_1m_2)\chi(-1)$$
$$\times \sum_{\substack{t \,(\mathrm{mod}\, q) \\ p|t}} (R_q(m_3t) - R_q(m_3))\chi(t + 1) \sideset{}{^*}\sum_{u \,(\mathrm{mod}\, q)} \overline{\chi}(rut + 1)\chi(1 - ru).$$

For the inner sum over u, apply $u \to u^{-1}$, giving

$$\sideset{}{^*}\sum_{u \,(\mathrm{mod}\, q)} \chi(1 - u^{-1}r)\overline{\chi}(1 + u^{-1}rt) = \sum_{u \,(\mathrm{mod}\, q)} \chi(u - r)\overline{\chi}(u + rt),$$

where we could omit the condition $(u, q) = 1$ since $p|t$. By Lemma 6.3, this equals $R_q(r(t + 1)) = R_q(r)$, provided $(t + 1, q) = 1$. Hence

$$S_2 = \chi_0(m_1 m_2)\chi(-1)R_q(r) \sum_{\substack{t \,(\mathrm{mod}\, q) \\ p|t}} \chi(t + 1)(R_q(m_3 t) - R_q(m_3)).$$

To complete the proof, we will show

$$\sum_{\substack{t \,(\mathrm{mod}\, q) \\ p|t}} \chi(t + 1)(R_q(m_3 t) - R_q(m_3)) = q\chi_0(m_3).$$

If $q = p$, this is immediate, noting $R_p(0) - R_p(m_3) = p\chi_0(m_3)$, so suppose $q = p^k$, $k \geq 2$. If $(p, m_3) = 1$, it is easy to verify the claim using the evaluation $R_q(m_3 t) = \sum_{d|(q,t)} d\mu(q/d)$ and Lemma 6.1. If $p|m_3$, then $R_q(m_3 t)$ is periodic (in t) of period p^{k-1}, so the sum vanishes by Lemma 6.1. $\qquad\square$

6.4. *The case* $q = p^k$, ψ *of conductor* p^j, $1 \leq j < k$.

Conjecture 6.6. Suppose χ has conductor p^k, and ψ has conductor p^j, with $1 \leq j < k$. Then

$$(6.4) \qquad \sum_{u,y \,(\mathrm{mod}\, p^j)} \psi(uy)\chi(1 + p^{k-j}y)\chi(1 - p^{k-j}u)\overline{\chi}(1 + uyp^{2(k-j)}) = O(p^j).$$

LEMMA 6.7. *Conjecture* 6.6 *holds in case* $k = 2, j = 1$.

Proof. By converting to additive characters (as in (6.2)), one may show $\sum_{x(\mathrm{mod}\, p)} \psi(x)\chi(1 + px) = \frac{\tau(\psi)\tau(\overline{\chi\psi})}{\tau(\overline{\chi})}$, which has absolute value $\sqrt{p}$. In the case $k = 2, j = 1$ the factor $\chi(1 + uyp^{2(k-j)})$ is identically 1, and so (6.4) is the product of two sums of this type. $\qquad\square$

In the following lemma and its proof, we use the convention that if χ is a Dirichlet character, $x \in \mathbb{Q}$, $x \notin \mathbb{Z}$, then $\chi(x) = 0$.

LEMMA 6.8. *Let* χ, ψ *be as in Conjecture* 6.6, *and suppose* (6.4) *holds. Then*

$$\widehat{H}(\psi, \chi, m_1, m_2, m_3, r)$$

$$(6.5) \qquad = \begin{cases} 0, & (m_1 m_2 m_3 r, p) = 1 \\ \chi_0(\frac{r}{p^{k-j}})O(p^{2k-\frac{j}{2}}), & p|r, \; m_1 m_2 m_3 = 1 \\ \chi_0(\frac{m_1}{p^{k-j}})\chi_0(\frac{m_2}{p^{k-j}})\chi_0(\frac{m_3}{p^{k-j}})O(p^{3k-\frac{3j}{2}}), & p|m_1 m_2 m_3, \; r = 1 \\ 0, & p|r, \; p|m_1 m_2 m_3. \end{cases}$$

In particular, the bound (6.5) *holds unconditionally for* $k = 2$, $j = 1$. *Furthermore, if in the second line of* (6.5) *the* $O(p^{2k-\frac{j}{2}})$ *is replaced by* $O(p^{2k+\frac{j}{2}})$ *and in the third line the* $O(p^{3k-\frac{3j}{2}})$ *is replaced by* $O(p^{3k-\frac{j}{2}})$, *then the revised bounds holds unconditionally for all* $1 \leq j < k$.

Proof. We begin with the observation

$$(6.6) \qquad \sum_{v \,(\mathrm{mod}\ q)} e_q(m_3 vt)\overline{\psi}(v) = p^{k-j}\tau(\overline{\psi})\psi\Big(\frac{m_3 t}{p^{k-j}}\Big).$$

Using (6.6) in (5.13), we have
(6.7)
$$\widehat{H} = p^{k-j}\tau(\overline{\psi}) \sum_{t,u \,(\mathrm{mod}\ p^k)} \chi(t + m_2 u)\overline{\chi}(rt + m_1 m_2)\overline{\chi}(u)\chi(-m_1 + ru)\psi\Big(\frac{m_3 t}{p^{k-j}}\Big).$$

First suppose that $(m_1 m_2 m_3 r, p) = 1$. Then changing variables $t \to ut$, we have

$$\widehat{H}(\psi,\chi,1,1,1,1) = p^{k-j}\tau(\overline{\psi}) \sum_{t,u \,(\mathrm{mod}\ p^k)} \chi(t+1)\overline{\chi}(ut+1)\chi(-1+u)\psi(u)\psi\Big(\frac{t}{p^{k-j}}\Big).$$

Note that $\overline{\chi}(ut + 1)\psi(u)$ is periodic in u of period p^j, since $p^{k-j}|t$ and ψ has conductor p^j. Hence by Lemma 6.1 the sum over u vanishes, as desired.

Now suppose $p|r$ and $m_1 m_2 m_3 = 1$. Then

$$\widehat{H}(\psi,\chi,1,1,1,r) = p^{k-j}\tau(\overline{\psi}) \sum_{t,u \,(\mathrm{mod}\ p^k)} \chi(t+u)\overline{\chi}(rt+1)\overline{\chi}(u)\chi(-1+ru)\psi\Big(\frac{t}{p^{k-j}}\Big).$$

Changing variables $t \to up^{k-j}y$ (where y now runs modulo p^j), we have

$$\widehat{H}(\psi,\chi,1,1,1,r)$$
$$= p^{k-j}\tau(\overline{\psi}) \sum_{y \,(\mathrm{mod}\ p^j)} \sideset{}{^*}\sum_{u \,(\mathrm{mod}\ p^k)} \chi(1 + p^{k-j}y)\chi(-1 + ru)\overline{\chi}(1 + rp^{k-j}uy)\psi(uy).$$

We claim the u-sum vanishes if $v_p(r) \neq k - j$, as we now show. Note that $\overline{\chi}(1 + rp^{k-j}uy)\psi(u)$ is periodic in u with period p^j, while if $v_p(r) < k-j$, then $\chi(-1 + ru)$ has period at least p^{j+1}. Lemma 6.1 then shows the claim. On the other hand, if $v_p(r) > k - j$, then $\chi(-1 + ru)\overline{\chi}(1 + rp^{k-j}uy)$ is periodic with period p^{j-1}, while $\psi(u)$ has least period p^j. Again, Lemma 6.1 shows the claim. Thus we may now restrict attention to $r = p^{k-j}$, in which case $\widehat{H}(\psi,\chi,1,1,1,p^{k-j})$ equals

$$p^{k-j}\tau(\overline{\psi}) \sum_{y \,(\mathrm{mod}\ p^j)} \sum_{u \,(\mathrm{mod}\ p^k)} \chi(1 + p^{k-j}y)\chi(-1 + p^{k-j}u)\overline{\chi}(1 + p^{2(k-j)}uy)\psi(uy).$$

The summand is periodic in u modulo p^j, so it is the same sum repeated p^{k-j} times. The conjectured bound (6.4) then finishes the job. Bounding the sum trivially gives an unconditional bound that is weaker by a factor p^j.

Now suppose $p|m_1m_2m_3$ and $r = 1$. We claim that $\widehat{H} = 0$ unless $p^{k-j}||m_i$ for each $i = 1, 2, 3$. By symmetry, we may assume $p|m_2$, say. Under this condition, the summand in (6.7) vanishes unless $(p, t) = 1$, in which case we must assume $p^{k-j}||m_3$. By symmetry again, this implies that the sum vanishes unless $p^{k-j}||m_1, m_2$ also. Then $\widehat{H}(\psi, \chi, p^{k-j}, p^{k-j}, p^{k-j}, 1)$ equals

$$\widehat{H} = p^{k-j}\tau(\overline{\psi}) \sum_{t,u\,(\mathrm{mod}\,p^k)} \chi(t + p^{k-j}u)\overline{\chi}(t + p^{2(k-j)})\overline{\chi}(u)\chi(-p^{k-j} + u)\psi(t).$$

Changing variables $t \to ut$, followed by $t \to t^{-1}$ and $u \to u^{-1}$, this becomes

$$\widehat{H} = p^{k-j}\tau(\overline{\psi}) \sum_{t,u\,(\mathrm{mod}\,p^k)} \chi(1 + p^{k-j}t)\chi(1 - p^{k-j}u)\overline{\chi}(1 + p^{2(k-j)}tu)\overline{\psi}(tu).$$

The summand is periodic modulo p^j, so it is the same sum repeated $p^{2(k-j)}$ times, and the conjectured bound (6.4) finishes the bound in this case. Bounding the sum trivially gives an unconditional bound that is weaker by a factor p^j.

Lastly, the case with $p|r$ and $p|m_1m_2m_3$ is covered by (5.15). $\qquad\square$

The most important case in the evaluation of $\widehat{H}$ occurs with (6.3), and it is crucial to have a strong bound on $g(\chi, \psi)$, which we claim with the following

THEOREM 6.9. *Let $g(\chi, \psi)$ be given by (1.9), where χ is primitive modulo q. For $q = p$ or $q = p^2$, we have*

$$|g(\chi, \psi)| \ll q.$$

We prove Theorem 6.9 in Section 9.

6.5. *Estimates for $H_\chi(m_1, m_2, m_3, r)$ in case some $m_j = 0$.* The calculations in this section may also be used to bound H_χ in case some $m_j = 0$, by way of (5.11) (of course, one could calculate H_χ directly). From Lemma 6.4 and the unconditional parts of Lemma 6.8, observe that $\widehat{H}(\psi, \chi, m_1, m_2, m_3, r) = 0$ if some $m_j = 0$, except in the case that ψ is the trivial character modulo q, in which case from Lemma 6.5 we deduce $|\widehat{H}(\psi, \chi, m_1, m_2, m_3, r)| \leq (m_1, q)(m_2, q)(m_3, q)$ by the trivial bound on the Ramanujan sums. Therefore by (5.11), we have

$$(6.8) \quad |H_\chi(m_1, m_2, m_3, r)| \ll q^{-1}(m_1, q)(m_2, q)(m_3, q)q^\varepsilon \quad \text{if } m_1m_2m_3 = 0.$$

It is useful to record that from (5.3), we deduce

$$(6.9) \quad |G(m_1, m_2, m_3, c)| \ll \frac{q^\varepsilon}{cq}\frac{(m_1, q)(m_2, q)(m_3, q)}{q} \quad \text{if } m_1m_2m_3 = 0.$$

7. Estimation of Z_{fin}

Let η_j, $j = 1, 2, 3, 4$ denote any unimodular completely multiplicative functions, and define

$$Z_{\text{fin},p}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sum_{\substack{a_1, a_2, a_3, d | p^\infty \\ (a_1, d) = 1}} \frac{\eta_1(a_1)\eta_2(a_2)\eta_3(a_3)\eta_4(d)}{a_1^{\sigma_1} a_2^{\sigma_2} a_3^{\sigma_3} d^{\sigma_4}} \widehat{H}(\psi_p, \chi_p, a_1, a_2, a_3, d).$$

LEMMA 7.1. *Let $Z_{\text{fin},p}$ be as above, with $q = p^k$, $k \geq 1$, and χ_p primitive modulo q. If ψ_p has conductor p^j with $1 \leq j < k$, assume Conjecture 6.6 holds for χ_p, ψ_p. If $\sigma_j \geq \sigma > 1/2$ for all j, then*

$$(7.1) \qquad Z_{\text{fin},p}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \ll_{\sigma,\varepsilon} \delta_\psi q^{1/2} |g(\chi, \psi)| + q^{3/2+\varepsilon},$$

where δ_ψ is the indicator function of the property that ψ is primitive (of conductor p^k). If $\sigma_j \geq \sigma > 1$ for all j, and ψ_p is the trivial character, then

$$(7.2) \qquad Z_{\text{fin},p}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \ll_{\sigma,\varepsilon} q^{1+\varepsilon}.$$

Remark. This result is unconditional for $k \leq 2$.

Proof. First suppose that ψ is primitive modulo q. By Lemma 6.4, all terms except $a_1 = a_2 = a_3 = d = 1$ vanish, giving the result.

Now suppose that ψ is the trivial character. By Lemma 6.5, we have

$$|Z_{\text{fin},p}| \leq q \sum_{r=0}^\infty \frac{(p^k, p^r)}{p^{r\sigma_4}} + \sum_{a_1, a_2, a_3 \geq 0} \frac{(p^k, p^{a_1})(p^k, p^{a_2})(p^k, p^{a_3})}{p^{a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3}}.$$

which is bounded consistently with the lemma.

Finally, consider ψ of conductor p^j, $1 \leq j < k$. Lemma 6.8, which depends on Conjecture 6.6, gives

$$|Z_{\text{fin},p}| \ll \frac{p^{2k-\frac{j}{2}}}{p^{(k-j)\sigma_4}} + \frac{p^{3k-\frac{3j}{2}}}{p^{(k-j)(\sigma_1+\sigma_2+\sigma_3)}} \ll p^{3k/2}.$$

This is consistent with (7.1). (Note that the bound (7.2) is not claimed in this case.) $\square$

8. Estimation of Z

8.1. *The main lemma.* Recall Z is given by (5.16).

LEMMA 8.1. *Suppose q is cube-free. There exists a decomposition $Z = Z_0 + Z_1$, where Z_0 and Z_1 satisfy the following properties. Firstly, Z_0 is meromorphic for $\text{Re}(s_j) \geq \sigma > 1/2$ for all j and analytic for $\text{Re}(s_j) \geq \sigma > 1$ for all j. It has a pole whenever some $s_j = 1$ and the other variables are fixed. In the region $\text{Re}(s_j) \geq \sigma > 1$ it satisfies the bound*

$$Z_0(s_1, s_2, s_3, s_4) \ll_{\sigma,\varepsilon} q^\varepsilon.$$

Secondly, Z_1 is analytic for $\mathrm{Re}(s_j) \geq \sigma \geq 1/2$ for all j, wherein it satisfies the bound

$$(8.1) \qquad \int_{-T}^{T} |Z_1(\sigma + it, \sigma + it, \sigma + it, \sigma - it)| dt \ll_\varepsilon q^{3/2+\varepsilon} T^{1+\varepsilon}$$

for $T \gg 1$. The same bound stated for Z_1 also hold for Z_0, provided $1/2 \leq \mathrm{Re}(s_j) \leq 0.99$.

Remark. The statement of Lemma 8.1 is essentially equivalent to [PY19b, Prop. 3].

Proof. Let Z_0 be the contribution to Z from the trivial character, and let $Z_1 = Z - Z_0$. All the desired estimates follow from the previous estimates on Z_{fin} and a bound on the fourth moment of Dirichlet L-functions (see [Pet15, Lem. 8] for instance). □

Conjecture 8.2. The statement of Lemma 8.1 holds for any q.

Remark. The proofs of the cubic moment bounds only need the properties of Z presented in Lemma 8.1. Therefore, if Conjecture 8.2 is true, then all the cubic moment bounds stated in the introduction of this paper are valid for arbitrary q.

9. Bounding $g(\chi, \psi)$: the proof of Theorem 6.9

9.1. *The case $q = p$.* In this subsection, we prove Theorem 6.9 in the case where $q = p$ is prime. Conrey and Iwaniec [CI00] proved $g(\chi, \psi) \ll p$ in the case that χ is the quadratic character. However, their proof does not seem to generalize: they conclude from Deligne's theorem that the bound $g(\chi, \psi) \ll p$ holds for all except at most one primitive ψ. The possible exceptional ψ can only be the quadratic character $\psi = \chi$, and then $g(\chi, \chi)$ has a special structure that Conrey and Iwaniec exploited to show $g(\chi, \chi) \ll p$ by elementary means. When χ is not quadratic, this special structure is not present, and it is not clear whether the bound $g(\chi, \psi) \ll p$ for ψ quadratic has an elementary proof.

To prove Theorem 6.9 we instead use Deligne's second proof of the Riemann Hypothesis [Del80]. We analyze the sum $g(\chi, \psi)$ by writing it as

$$\sum_x \chi(x)\overline{\chi}(x+1) \left(\sum_y \overline{\chi}(y)\chi(y+1)\psi(xy-1) \right)$$

and realize the inner sum as a trace function in x of a sheaf $\mathcal{G}$. The sheaf $\mathcal{G}$ will then be compared with the sheaf that corresponds to the trace function $x \mapsto \chi(x)\overline{\chi}(x+1)$ to show cancellation in both variables. In executing this strategy, we have benefited greatly from the recent works of Fouvry, Kowalski

and Michel, which have served to make the theorems of Deligne and Katz on trace functions more amenable to analytic applications.

Proof. Suppose χ and ψ are primitive modulo p, and let χ_m, ψ_m be the characters derived from χ, ψ by composing with the norm map $N : \mathbb{F}_{p^m} \to \mathbb{F}_p$. Let

$$g(\chi_m, \psi_m) = \sum_{x,y \in \mathbb{F}_{p^m}} \chi_m(x)\overline{\chi_m}(x+1)\overline{\chi_m}(y)\chi_m(y+1)\psi_m(xy-1).$$

By the Grothendieck-Lefschetz trace formula [Del77, Rapport, Thm. 3.2] and the Riemann hypothesis of Deligne [Del80], we have that there exist algebraic numbers $\alpha_{i,+}$ and $\alpha_{i,-}$ with $|\alpha_{i,+}| = p^{k_i/2}$, $|\alpha_{i,-}| = p^{\ell_i/2}$ with $k_i, \ell_i \in \mathbb{Z}$ such that

$$(9.1) \qquad g(\chi_m, \psi_m) = -\sum_{i=1}^{N_+} \alpha_{i,+}^m + \sum_{i=1}^{N_-} \alpha_{i,-}^m.$$

Results of Adolphson-Sperber or Katz [Kat01, Thm. 12] show that $N_+, N_- \ll 1$, independently of χ, ψ, p. Thus, to prove Theorem 6.9 in the case that χ, ψ are primitive modulo p, it suffices to show that $|\alpha_{i,+}|, |\alpha_{i,-}| \leq p$.

We show that $|\alpha_{i,+}|, |\alpha_{i,-}| \leq p$ using the theory of ℓ-adic sheaves and trace functions; for background, see [Del77], [Kat88], [FKMS19]. Let ℓ be a prime distinct from p, and let $\iota : \overline{\mathbb{Q}}_\ell \to \mathbb{C}$ be a fixed isomorphism. If X is an algebraic variety over $\mathbb{F}_p$, then by "sheaf" or "ℓ-adic sheaf" we will mean a constructible $\overline{\mathbb{Q}}_\ell$-sheaf on X. Note that ℓ is always assumed distinct from the base field of X. If $\mathcal{F}$ is a sheaf on X and $\overline{x} \in X(\overline{\mathbb{F}}_p)$ is a geometric point of X, then we write $\mathcal{F}_{\overline{x}}$ for the stalk of $\mathcal{F}$ at $\overline{x}$.

For any ℓ-adic sheaf $\mathcal{F}$ on X, its trace function $t_\mathcal{F}(x)$ is defined to be the value at $x \in X(\mathbb{F}_p)$ of the trace of the Frobenius endomorphism of $\mathbb{F}_p$ acting on $\mathcal{F}_x$. That is,

$$t_\mathcal{F}(x) = \iota((\mathrm{Tr}\mathcal{F})(\mathbb{F}_p, x)) = \iota(\mathrm{Tr}(\mathrm{Fr}_p|\mathcal{F}_x)).$$

Let

$$\mathcal{F}_1 = \mathcal{L}_{\chi((Y+1)Y^{-1})}$$

be the Kummer sheaf attached to the character $\chi\left(\frac{Y+1}{Y}\right)$, whose trace function is $\overline{\chi}(y)\chi(y+1)$. Thus $\mathcal{F}_1$ is an ℓ-adic sheaf on $\mathbb{A}^1$; it is a middle-extension sheaf, pure of weight 0, and lisse on $\mathbb{A}^1 - \{0, -1\}$. It is of rank 1, hence geometrically irreducible. Let

$$\mathcal{K} = \mathcal{L}_{\psi(XY-1)}$$

be the middle extension of the Kummer sheaf attached to $\psi(XY-1)$ on $\mathbb{A}^1 \times \mathbb{A}^1$. Let $Z \subset \mathbb{A}^2$ be the closed set defined by the equation $XY = 1$. The sheaf $\mathcal{K}$ is lisse, of rank 1 and pure of weight 0 on the dense open set $V = \mathbb{A}^2 - Z$. Since

ψ is non-trivial, the middle extension sheaf is identical to the extension by 0 of $\mathcal{K}$ restricted to V.

Let $p_j : \mathbb{A}^1 \times \mathbb{A}^1 \to \mathbb{A}^1$, $j = 1, 2$ be the two canonical projections, and let

$$\mathcal{H} = p_2^* \mathcal{F}_1 \otimes \mathcal{K}.$$

The sheaf $\mathcal{H}$ is lisse on the dense open set $U = \mathbb{A}^2 - D$, where $D \subset \mathbb{A}^2$ is the divisor

$$D = Z \cup \mathbb{A}^1 \times \{0\} \cup \mathbb{A}^1 \times \{-1\}.$$

For $i = 0, 1, 2$, define the ℓ-adic sheaves

$$T_{\mathcal{K}}^i(\mathcal{F}_1) := R^i p_{1,!}(\mathcal{H}),$$

where $R^i p_{1,!}$ is the higher direct image with compact supports. The sheaf $\mathcal{G} = T_{\mathcal{K}}^1(\mathcal{F}_1)$ is is the "cohomological transform" of $\mathcal{F}_1$ defined by the "kernel" $\mathcal{K}$, in the sense of Katz's affine cohomological transforms and of [FKM21].

LEMMA 9.1. *If χ and ψ are non-trivial Dirichlet characters modulo p, then*

$$t_{\mathcal{G}}(x) = -\sum_{y \in \mathbb{F}_p} t_{\mathcal{F}_1}(y) t_{\mathcal{K}}(x, y).$$

Proof. Let $U_x = U \cap \{x\} \times \mathbb{A}^1$ be the open set on which $\mathcal{H}$ restricted to $\{x\} \times \mathbb{A}^1$ is lisse. Precisely, we have $U_x = \{x\} \times (\mathbb{A}^1 - \{0, -1, 1/x\})$. (Below we take restrictions of $p_2^* \mathcal{F}_1$ and $\mathcal{K}$ to $\{x\} \times \mathbb{A}^1$ without mention.)

There are three representations of $\mathrm{Gal}(\overline{\mathbb{F}}_p / \mathbb{F}_p)$ given by $H_c^i(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H})$ for $i = 0, 1, 2$. The Grothendieck-Lefschetz trace formula asserts that

$$\sum_{y \in U_x(\mathbb{F}_p)} t_{\mathcal{H}}(x, y) = \mathrm{Tr}(\mathrm{Fr}_p | H_c^0(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H}))$$

$$- \mathrm{Tr}(\mathrm{Fr}_p | H_c^1(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H})) + \mathrm{Tr}(\mathrm{Fr}_p | H_c^2(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H})),$$

where $\mathrm{Fr}_p \in \mathrm{Gal}(\overline{\mathbb{F}}_p / \mathbb{F}_p)$ is the Frobenius automorphism. By standard operations with Galois representations, and the fact that $\mathcal{F}_1$ and $\mathcal{K}$ are extension by 0 sheaves, we have that

$$\sum_{y \in \mathbb{F}_p} t_{\mathcal{F}}(y) t_{\mathcal{K}}(x, y) = \sum_{y \in U_x(\mathbb{F}_p)} t_{\mathcal{H}}(x, y).$$

Furthermore, by the proper base change theorem (see [Del77, Arcata, IV, Th. 5.4]) we have that $H_c^i(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H})$ is naturally isomorphic to the stalk at x of $T_{\mathcal{K}}^i(\mathcal{F}_1)$. Therefore, to prove the lemma, it suffices to show that all of the stalks of $T_{\mathcal{K}}^0(\mathcal{F}_1)$ and $T_{\mathcal{K}}^2(\mathcal{F}_1)$ are 0.

First we show that the stalks of $T_{\mathcal{K}}^0(\mathcal{F}_1)$ are all zero. We claim that $H_c^0(U_{x, \overline{\mathbb{F}}_p}, \mathcal{H}) = 0$, and so $T_{\mathcal{K}}^0(\mathcal{F}_1) = 0$ as well. Since χ is non-trivial, $\mathcal{F}_1$ is a middle extension sheaf, and so is $p_2^* \mathcal{F}_1$. Since both $p_2^* \mathcal{F}_1$ and $\mathcal{K}$ are middle-extension, we have, e.g. by [FKM21, Lem. 4.2], that $H_c^0(\{x\} \times \mathbb{A}^1, \mathcal{H}) = 0$. Let

$\pi : (\{x\} \times \mathbb{A}^1) - U_x \to \operatorname{Spec} \overline{\mathbb{F}}_p$ be the structure morphism. The sheaf $R^{-1}\pi_! \mathcal{H}$ vanishes by definition, so $H_c^{-1}((\{x\} \times \mathbb{A}^1) - U_x, \mathcal{H}) = 0$. By excision (see [Del77, Sommes Trig. $(2.5.1)^*$]) and the vanishing of the above two cohomology groups, we have that $H_c^0(U_{x,\overline{\mathbb{F}}_p}, \mathcal{H}) = 0$ as well.

Now we show that the stalks of $T_{\mathcal{K}}^2(\mathcal{F}_1)$ are all zero. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are any two geometrically irreducible sheaves, lisse on U_x, then $H_c^2(U_{x,\overline{\mathbb{F}}_p}, \mathcal{L}_1 \otimes \mathcal{L}_2) \neq 0$ if and only if $\mathcal{L}_1 \simeq D(\mathcal{L}_2)$ on a dense open set where both sheaves are lisse, as one can see by the co-invariants formula (see [Del80, $(1.4.1)$b]) and Schur's lemma. In our case, it suffices to consider the $G^{\mathrm{geom}} = \operatorname{Gal}(\overline{\mathbb{F}_p(X)}/\overline{\mathbb{F}}_p(X))$-invariants acting on the stalk of $p_2^* \mathcal{F}_1$ and $\mathcal{K}$ at a lisse geometric point. Since χ is non-trivial (this is crucial), we have that $\mathcal{F}_1$ is ramified at 0 whereas $\mathcal{K}$ is not. Therefore the inertia group at zero $I_0 \subset G^{\mathrm{geom}}$ acts non-trivially on the stalk of $\mathcal{F}_1$ at any lisse point, whereas I_0 acts trivially on any stalk of $\mathcal{K}$. Therefore the two sheaves cannot be geometrically isomorphic, and so the H_c^2 vanishes. $\qquad\square$

By Lemma 9.1 and the fact that $\mathcal{F}_1$ is middle extension, we have

$$(9.2) \qquad g(\chi, \psi) = -\sum_{u \in \mathbb{F}_p} \overline{t_{\mathcal{F}_1}(u)} t_{\mathcal{G}}(u).$$

By the Riemann hypothesis of Deligne [Del80, Thm. 3.3.1], $\mathcal{G}$ is mixed of weights ≤ 1, so to apply the orthogonality form of the Riemann hypothesis (e.g., [FKMS19, Thm. 5.2]), we would need to show that the part of weight 1 of $\mathcal{G}$, say $\mathcal{G}_0$, is geometrically irreducible and not geometrically isomorphic to $\mathcal{F}_1$. It is not difficult to see that $\mathcal{G}$ has generic rank 2, and we would like to argue that this prevents $\mathcal{G}_0$ from being geometrically isomorphic to $\mathcal{F}_1$. However, it is less clear that $\mathcal{G}_0$ itself has generic rank 2.

Recall $U = \mathbb{A}^2 - D$, and let j be the open embedding of U in $\mathbb{A}^2$. To handle the issue raised in the previous paragraph, let us introduce the modified sheaf

$$\widetilde{\mathcal{H}} := j_!(j^*\mathcal{H})$$

and the corresponding cohomological transform sheaves

$$\widetilde{T}_{\mathcal{K}}^i(\mathcal{F}_1) := R^i p_{1,!}(\widetilde{\mathcal{H}}).$$

We have defined $\widetilde{\mathcal{H}}$ in order that it satisfy the hypotheses of Deligne's semicontinuity theorem [Lau81], used in the proof of part (1) of the following lemma.

LEMMA 9.2. *Suppose that χ and ψ are non-trivial modulo p. The sheaf $\widetilde{\mathcal{G}} = \widetilde{T}_{\mathcal{K}}^1(\mathcal{F}_1)$*

(1) *is lisse on the dense open set $W = \mathbb{A}^1 - \{0, -1\}$,*

(2) *is geometrically irreducible and pure of weight 1 on W, and*

(3) *has generic rank 2.*

Proof. (1) In order to prove that $\widetilde{\mathcal{G}}$ is lisse on W, we will use Deligne's semicontinuity theorem [Lau81, Cor. 2.1.2]. Consider

$$\widetilde{p}_1 : \mathbb{A}^1 \times \mathbb{P}^1 \to \mathbb{A}^1,$$

which is a smooth and proper morphism of relative dimension 1. Abusing notation, we continue to write $\widetilde{\mathcal{H}}$ for the extension by 0 of $\widetilde{\mathcal{H}}$ from $\mathbb{A}^1 \times \mathbb{A}^1$ to $\mathbb{A}^1 \times \mathbb{P}^1$. By definition, we have $\widetilde{\mathcal{G}} = R^1_{\widetilde{p}_1,*} \widetilde{\mathcal{H}}$.

Let $\widetilde{D}$ be the complement in $\mathbb{A}^1 \times \mathbb{P}^1$ of the open set U, that is $\widetilde{D} = D \cup \mathbb{A}^1 \times \{\infty\}$. By restriction, $\widetilde{p}_1$ defines a proper smooth morphism of relative dimension 1

$$X \to W = \mathbb{A}^1 - \{0, -1\},$$

where $X = \widetilde{p}_1^{-1}(W)$. The intersection $\widetilde{D} \cap X$ is a divisor in X, which is flat and finite (of degree 4) over W. The sheaf $\widetilde{\mathcal{G}}$ is lisse on the complement of $\widetilde{D} \cap X$ in $\mathbb{A}^1 \times \mathbb{P}^1$.

Let $x \in W$. We identify the fiber C_x of $\widetilde{p}_1$ over x with $\mathbb{P}^1$, and so the restriction of $\widetilde{\mathcal{H}}$ to C_x is identified with a lisse sheaf on the dense open set (abusing notation)

$$U_x = \mathbb{A}^1 - \{0, -1, 1/x, \infty\} \subset \mathbb{P}^1.$$

The restriction of the sheaf $\widetilde{\mathcal{H}}$ to C_x is at most tamely unramified everywhere, hence the function φ of [Lau81, Thm. 2.1.1] is constant equal to 0 on points of W. Then we have by Corollary 2.1.2 of loc. cit. that $R^1 \widetilde{p}_{1,*} \widetilde{\mathcal{H}}$ is lisse on W.

(2) The sheaf $\widetilde{\mathcal{G}}$ is mixed of weights ≤ 1 on W by the Riemann hypothesis of Deligne [Del80, Thm. 3.3.1]. Furthermore, the part of $\widetilde{\mathcal{G}}$ of weight 1 is geometrically irreducible on W by the diophantine criterion for geometric irreducibility. Indeed, $\mathcal{F}_1$ is not geometrically isomorphic to the Kummer sheaf $\mathcal{L}_\psi$ attached to ψ, since $\mathcal{F}_1$ is ramified at -1 and $\mathcal{L}_\psi$ is not, and so the hypothesis of [FKM21, Prop. 5.12 and Rem. 5.14] is satisfied.

Finally, we prove that $\widetilde{\mathcal{G}}$ is pure of weight 1 on W by applying [KMS17, Lem. 4.22(2)] to the morphism $\widetilde{p}_1 : X \to W$ and the sheaf $\widetilde{\mathcal{H}}$ on X. For any $x \in \mathbb{P}^1$, the sheaf pullback $\widetilde{\mathcal{H}}_x$ to $C_x = \{x\} \times \mathbb{P}^1$ has no punctual section nor trivial subrepresentation (as a lisse sheaf on U_x). Thus the assumptions of loc. cit. are satisfied.

It follows that for any $x \in W$, the part of weight < 1 of the stalk at x of $\widetilde{\mathcal{G}} = R^1 \widetilde{p}_{1,*} \widetilde{\mathcal{H}}$ is isomorphic to

$$\bigoplus_{y \in C_x - U_x} (\widetilde{\mathcal{H}}_x)^{I_y}_{\overline{\eta}} / (\widetilde{\mathcal{H}}_x)_{\overline{y}}.$$

But we already have $(\widetilde{\mathcal{H}}_x)^{I_y}_{\overline{\eta}} = 0$ at all singular points $y \in \{0, -1, 1/x, \infty\}$, so the direct sum vanishes. From Deligne's Theorem, we conclude that $\widetilde{\mathcal{G}}$ is pure of weight 1 on W.

(3) The stalk of $\widetilde{\mathcal{G}}$ over $x \in \mathbb{F}_p$ is $H^1_c(\{x\} \times \mathbb{A}^1, \widetilde{\mathcal{H}})$. By the Euler-Poincaré formula [Kat88, 8.5.2, 8.5.3], if $x \neq -1$ then the dimension of this cohomology group is $-1 + 3 = 2$ for the three tamely ramified points $0, -1, 1/x$ of $\widetilde{\mathcal{H}}$. Hence the generic rank is 2. $\qquad\qquad\square$

To compare $\widetilde{\mathcal{H}}$ with $\mathcal{H}$, observe that the stalks of $\widetilde{\mathcal{H}}$ are equal to those of $\mathcal{H}$ outside D, and are 0 along D. Thus, the stalks of $\mathcal{H}$ and $\widetilde{\mathcal{H}}$ may only differ on $\widetilde{D}$ (and this can only happen if $\chi = \overline{\psi}$). Indeed, in the case $\chi = \overline{\psi}$ note that $t_{\mathcal{H}}(-1, -1) = \overline{\chi}(-1)^2$, whereas $t_{\widetilde{\mathcal{H}}}(-1, -1) = 0$. As an aside, one wonders whether $g(\chi, \overline{\chi})$ admits a "special structure" for χ complex that could be used to give a simpler the proof of the bound $g(\chi, \overline{\chi}) \ll p$ in that special case, as was exploited by Conrey and Iwaniec in the case that χ is quadratic [CI00, §14].

By the discussion in the preceding paragraph, we have that $t_{\mathcal{G}}(u) = t_{\widetilde{\mathcal{G}}}(u) + O(1)$. Thus, setting

$$s(\chi, \psi) = - \sum_{\mathbb{F}_p - \{0, -1\}} \chi(x)\overline{\chi}(x+1) \sum_y \overline{\chi}(y)\chi(y+1)\psi(xy - 1),$$

we have

$$(9.3) \qquad g(\chi, \psi) = -s(\chi, \psi) + O(p) = - \sum_{x \in W(\mathbb{F}_p)} \overline{t_{\mathcal{F}_1}(x)} t_{\widetilde{\mathcal{G}}}(x) + O(p).$$

Since $\mathcal{F}_1$ is pure of weight 0 and rank 1 on W and $\widetilde{\mathcal{G}}$ is pure of weight 1 and geometrically irreducible of rank 2 on W by Lemma 9.2, $\widetilde{\mathcal{G}}$ cannot be isomorphic to $\mathcal{F}_1$. Thus, the Grothendieck-Lefschetz trace formula, the co-invariants formula [Del80, (1.4.1)b], and the Riemann hypothesis of Deligne imply that there exist algebraic numbers $\beta_{i,+}$ and $\beta_{i,-}$, with $|\beta_{i,+}| \leq p$, $|\beta_{i,-}| \leq p$ such that

$$(9.4) \qquad s(\chi_m, \psi_m) = \sum_{x \in W(\mathbb{F}_{p^m})} \overline{t_{\mathcal{F}_1}(N(x))} t_{\widetilde{\mathcal{G}}}(N(x)) = -\sum_{i=1}^{M_+} \beta_{i,+}^m + \sum_{i=1}^{M_-} \beta_{i,-}^m.$$

Here it is *not* clear that M_+ and M_- are bounded independently of χ, ψ, p. However, we can avoid this issue by appealing to the two-dimensional Riemann hypothesis of Deligne (9.1), in which situation we know that $N_+, N_- \ll 1$. A slight variation of [CI00, Lem. 13.2] shows that $|\alpha_{i,+}|, |\alpha_{i,-}| \leq p^{3/2}$, and we would like to show in fact that $\alpha_{i,+}$ and $\alpha_{i,i}$ are bounded by p. Suppose not. Then we would have

$$\limsup_{m \to \infty} \frac{|g(\chi_m, \psi_m)|}{p^{3m/2}} > 0.$$

But this is impossible by (9.4) since $|\beta_{i,+}|, |\beta_{i,-}| \leq p$. Thus $|\alpha_{i,+}|, |\alpha_{i,-}| \leq p$, so by (9.1) and the fact that $N_+, N_- \ll 1$ we have $g(\chi, \psi) \ll p$ for all χ, ψ primitive.

If ψ is not primitive, it must be the trivial character ψ_0, in which case we have $g(\chi, \psi_0) \ll p$ by Lemma 6.3, which completes the proof of Theorem 6.9 when $q = p$. $\qquad\square$

Remark. A proof of a minor variant of Theorem 6.9 also appears as [FKM21, Thm. 5.8], from which we drew inspiration in giving the proof that appears above. However, our proof departs from that of loc. cit. in that we have completely avoided the difficult main Theorems 2.3, 2.5, and 5.11 of [FKM21] on the behavior of the conductor under cohomological transforms.

9.2. *The case* $q = p^2$. This case can be treated by elementary means. Since χ is a Dirichlet character modulo p^2, the function $t \mapsto \chi(1 + pt)$ is an additive character on $\mathbb{Z}/p\mathbb{Z}$, so there exists an integer ℓ_χ so that

$$\chi(1 + pt) = e_p(\ell_\chi t).$$

Note that χ has conductor p^2 if and only if $(\ell_\chi, p) = 1$. Hence if a, b are integers with $(a, p) = 1$, then

(9.5) $$\chi(a + pb) = \chi(a)\chi(1 + p\overline{a}b) = \chi(a)e_p(\ell_\chi\overline{a}b),$$

where $\overline{a} \in \mathbb{Z}$ satisfies $a\overline{a} \equiv 1 \pmod{p^2}$. Now, for each $t, u \pmod{q}$, choose $a, b, c, d \in \mathbb{Z}$ such that $a + pb \equiv t \pmod{q}$ and $c + pd \equiv u \pmod{q}$. We have

$$\psi(ut - 1) = \psi(-1 + ac + p(bc + ad)) = \psi(-1 + ac)e_p(\ell_\psi(bc + ad)\overline{(-1 + ac)}).$$

Note that as t, u run through $\mathbb{Z}/q\mathbb{Z}$, each of the integers a, b, c, d represent every residue class modulo p. We obtain

(9.6)
$$g(\chi, \psi) = \sum_{a,c} \chi(a)\overline{\chi}(a + 1)\overline{\chi}(c)\chi(c + 1)\psi(-1 + ac) \sum_{b,d} e_p(\ell_\chi b(\overline{a} - \overline{(a + 1)}))$$
$$+ \ell_\psi bc\overline{(-1 + ac)})e_p(-\ell_\chi d(\overline{c} - \overline{(c + 1)}) + \ell_\psi ad\overline{(-1 + ac)}).$$

In particular, we have $(a(a + 1)c(c + 1)(ac - 1), p) = 1$ for every non-zero term of the sum over a and c, so all inversions modulo p here and below are justified. The sum over b equals p provided

(9.7) $$\ell_\chi(\overline{a} - \overline{(a + 1)}) \equiv -\ell_\psi c\overline{(-1 + ac)} \pmod{p}$$

and vanishes otherwise. Similarly, the sum over d equals p provided

(9.8) $$\ell_\chi(\overline{c} - \overline{(c + 1)}) \equiv \ell_\psi a\overline{(-1 + ac)} \pmod{p}$$

and vanishes otherwise. We claim that there at most two solutions to (9.7) and (9.8), whence

$$|g(\chi, \psi)| \leq 2q$$

for $q = p^2$. Along the way, we will also see that $g(\chi, \psi) = 0$ if ψ is not primitive.

Indeed, multiplying the first congruence by $a(a + 1)$ and the second one by $c(c + 1)$, we obtain the equivalent system

$$\ell_\chi \equiv -\ell_\psi ac(a+1)\overline{(-1 + ac)} \pmod{p}, \quad \ell_\chi \equiv \ell_\psi ac(c+1)\overline{(-1 + ac)} \pmod{p}.$$

Since $(\ell_\chi, p) = 1$, this implies that $g(\chi, \psi) = 0$ unless $(\ell_\psi, p) = 1$, which means ψ is primitive. Furthermore, we deduce that $a(a + 1)c \equiv -ac(c + 1) \pmod{p}$, whence $c \equiv -2 - a \pmod{p}$, which uniquely determines c in terms of a. Then we see that a must satisfy

$$(9.9) \qquad a(a + 1)(a + 2)\overline{(-1 + a(-2 - a))} \equiv \overline{\ell_\psi}\ell_\chi \pmod{p}.$$

Setting $A = \overline{\ell_\psi}\ell_\chi$, (9.9) is equivalent to

$$a(a + 2) \equiv -A(a + 1) \pmod{p}.$$

Hence a satisfies a certain monic quadratic polynomial, having at most two solutions modulo p. This gives the desired bound on g, completing the proof of Theorem 6.9.

10. **Archimedean aspects, part 1**

In this section, we derive the analytic properties of the weight functions $J_0^\pm$ and $J^\pm$ defined in (4.6) and (4.11).

10.1. *Approximate functional equations.* Recall from (4.3) the functions $V_j(y, t)$.

LEMMA 10.1. *For each $j = 1, 2$, $V_j(y, t)$ is an entire, even function in t for any given $y > 0$. Moreover, for $t \in \mathbb{R}$, it satisfies the bound*

$$(10.1) \qquad y^k(1/2 + it)^\ell \frac{\partial^{k+\ell}}{\partial y^k \partial t^\ell} V_j(y, t) \ll_{A,k,\ell} \left(1 + \frac{y}{1 + |t|^j}\right)^{-A}$$

for any $A > 0$. For $t = -i/2 + v$ with $v \in \mathbb{R}$, we have, for any $A > 0$,

$$(10.2) \qquad y^k \frac{\partial^k}{\partial y^k} V_j(y, -\tfrac{i}{2} + v) \ll_{A,k} \left(1 + \frac{y}{1 + |v|^j}\right)^{-A}.$$

Proof. By shifting the contour far to the right, we see that $V_j(y, t)$ is analytic for t in any fixed horizontal strip, so it can be extended as an entire function of t. It is clearly invariant under $t \to -t$.

Now assume $t \in \mathbb{R}$. First we show (10.1) in the case $k = \ell = 0$. We assume $\delta = 0$ for notational simplicity, the $\delta = 1$ case being similar. Stirling's asymptotic expansion gives that $\log \Gamma(z) = (z - \frac{1}{2}) \log(z) - z + \sum_{j=0}^{N} c_j z^{-j} +$

$O(|z|^{-N-1})$, for certain constants c_j, valid for $\mathrm{Re}(z) \geq 1/100$, say. From this
we deduce that if $|a|^2 \leq |z|$, $\mathrm{Re}(z) \geq 1/4$, then

$$(10.3) \qquad \log \frac{\Gamma(z+a)}{\Gamma(z)} = a \log z + \sum_{j=1}^{N} \frac{P_j(a)}{z^j} + O\Big(\frac{(1+|a|)^{2N+2}}{|z|^{N+1}}\Big)$$

for certain polynomials P_j of degree at most $2j$. Fix $\sigma \in \mathbb{R}$ so that $1/2+\sigma > 0$.
Then for $\mathrm{Re}(s) = \sigma$ and $|\mathrm{Im}(s)| \leq (1+|t|)^{1/4}$, we derive from (10.3) that

$$(10.4)$$
$$\exp(s^2)\frac{\Gamma_{\mathbb{R}}(1/2+s+it)}{\Gamma_{\mathbb{R}}(1/2+it)}$$
$$= \Big(\frac{|t|}{\pi}\Big)^{s/2} \exp(s^2)\Big(1 + \sum_{j=1}^{N} \frac{P_j(s)}{(1/2+it)^j} + O_{\sigma,N}((1+|t|)^{-\frac{N+1}{2}})\Big),$$

provided t is sufficiently large, and where P_j is a different collection of polynomials of degree $\leq 2j$. If $|\mathrm{Im}(s)| > (1+|t|)^{1/4}$, then a crude application of Stirling gives

$$\exp(s^2)\frac{\Gamma_{\mathbb{R}}(1/2+s+it)}{\Gamma_{\mathbb{R}}(1/2+it)} = O((1+|t|)^{\sigma/2}\exp(-\mathrm{Im}(s)^2/2)).$$

In any event, we shift the contour to $\mathrm{Re}(s) = A$ to see that $V_j(y,t) \ll_A$
$(1+|t|^j)^A y^{-A}$ for $y > 1 + |t|^j$. If $y \leq 1 + |t|^j$, we instead move the contour to
$\sigma = -1/4$, say. Accounting for the pole and bounding the integral on the new
line, we obtain $V_j(y,t) \ll 1$ in this case.

Next we consider derivatives. Note that differentiation k times with respect to y followed by multiplication by y^k gives an integral of the form (4.3)
back, but with $G_j(s)$ multiplied by a polynomial in s. The exponential decay
of $G_j(s)$ easily accommodates for this, showing (10.1) for $\ell = 0$, and any $k \geq 0$.
Differentiation of Stirling's formula with respect to t leads to (10.1) for any k, ℓ.

Next consider the case $t = -i/2 + v$ with $v \in \mathbb{R}$, so $it = 1/2 + iv$. For $y >
1+|t|^j$, we move the contour far to the right and bound it the same way. For $y \leq
1+|t|^j$, we shift left, to $-1/4$ again. We pass poles at $s = 0$ (as before) giving a
residue of 1, and at $s = -1/2 + iy = iv$. This latter residue is $O((1+|v|)^{-100})$,
i.e., uniformly bounded for $v \in \mathbb{R}$, using that the apparent pole of $\frac{1}{iv}$ at $v = 0$
is cancelled by a zero of $1/\Gamma_{\mathbb{R}}(-iv)$. It is not hard to see that (10.2) holds. $\square$

10.2. *Properties of J^+*. We invite the reader to recall the definition of
inert functions from Section 3.

LEMMA 10.2. *Let $J^+(x,\cdot)$ be defined as in* (4.11). *Then*

$$(10.5) \qquad \frac{\partial^k}{\partial x^k} J^+(x,\cdot) \ll_k x(x^{-k} + x^k)T^{k+1},$$

and $J^+(x,\cdot)$ is 1-inert with respect to the variables $t_1, t_2, t_3, c, m_1, m_2, m_3$.

We will use this for the relatively small values $x \ll T^{2+\varepsilon}$. In the complementary range, we have the following.

LEMMA 10.3. *Suppose for some $\varepsilon > 0$ that $1 \leq T^{2+\varepsilon} \ll x$. Then for any $A > 0$,*

$$J^+(x, \cdot) = \sum_{\pm} T^2 x^{-1/2} e^{\pm ix} W_\pm(x, \cdot) + O_{\varepsilon, A}(x^{-A}),$$

where $W_\pm(x, \cdot)$ is a function (depending additionally on ε, T, A) satisfying $x^k \frac{\partial^k}{\partial x^k} W_\pm(x, \cdot) \ll 1$. Moreover, $W_\pm(x, \cdot)$ is 1-inert with respect to the variables $t_1, t_2, t_3, m_1, m_2, m_3, c$.

Proof of Lemma 10.2. First consider the case $k = 0$. In (4.11) we shift the contour to the line $\mathrm{Im}(t) = -1/2$. Then from (10.2), and using $|\cosh(-\frac{\pi i}{2} + \pi y)| = |\sinh(\pi y)|$, we have

$$|J^+(x, \cdot)| \ll \int_{-\infty}^{\infty} \frac{|J_{1+2iy}(x)|}{|\sinh(\pi y)|} \frac{|y|(1 + y^2)}{T^2} \exp(-(y/T)^2) dy.$$

Next we claim that for any integer $a \geq 0$, we have

(10.6) $$\frac{|J_{1+a+2iy}(x)|}{|\sinh(\pi y)|} \ll \frac{1 + |y|}{|y|} \frac{(x/2)^{1+a}}{|1/2 + 2iy|^{a+1}}.$$

This bound can be derived with a little work from [GR00, 8.411.4] and Stirling's approximation. Taking $a = 0$, this implies (10.5) for $k = 0$.

We next extend this to $k \geq 1$. By [GR00, 8.472.2, 8.486.13] we have

(10.7) $$\frac{d}{dx} Z_\nu(x) = \frac{\nu}{x} Z_\nu(x) - Z_{\nu+1}(x),$$

valid for $Z_\nu = J_\nu$ as well as $Z_\nu = K_\nu$. Iterating this, we derive

(10.8) $$\frac{d^k}{dx^k} J_\nu(x) = \sum_{m=0}^{k} \frac{P_{k,m}(\nu)}{x^m} J_{\nu+k-m}(x),$$

where $P_{k,m}$ is a polynomial of degree at most m. By (10.8) and (10.6), we deduce that

$$\frac{|\frac{d^k}{dx^k} J_{1+2iy}(x)|}{|\sinh(\pi y)|} \ll_k \frac{1 + |y|}{|y|} \sum_{m=0}^{k} \frac{(1 + |y|)^m}{x^m} \frac{x^{1+k-m}}{(1 + |y|)^{1+k-m}}$$

$$\ll \frac{x}{|y|} \left(\frac{x^k}{(1 + |y|)^k} + \frac{(1 + |y|)^k}{x^k} \right).$$

It is then straightforward to derive (10.5) for all k.

The final statement of the lemma, that $J^+(x, \cdot)$ is 1-inert with respect to the other variables, follows from Lemma 10.1, since the only dependence of J^+ on these auxiliary parameters is via the factors $V_1(y_1, t) V_2(y_2, t)$ and the inert function w. $\quad\square$

Proof of Lemma 10.3. By [GR00, 8.411.11] and an interchange of orders of integration justified by integration by parts and Fubini, there exists an integral representation in the form

$$J^+(x, \cdot) = w(\cdot)T^2 \int_0^\infty \cos(x \cosh(v))g(v, \cdot)dv,$$

where

$$g(v, \cdot) = T^{-2} \int_{-\infty}^\infty t \tanh(\pi t) \frac{t^2 + \frac{1}{4}}{T^2} \exp(-(t/T)^2) \cos(2tv)V_1(\cdot, t)V_2(\cdot, t)\, dt.$$

Here g is a Schwartz-class function; more precisely, it satisfies the bounds

(10.9) $g^{(j)}(v, \cdot) \ll_{A,j} T^j(1 + |v|)^{-A}$ for all $A > 0$, $j \geq 0$

and is 1-inert with respect to the other variables by Lemma 10.1. Hence

$$J^+(x, \cdot) = \sum_\pm T^2 \int_0^\infty e^{\pm ix \cosh(v)}g(v, \cdot)dv = \sum_\pm T^2 e^{\pm ix}F_\pm(x, \cdot),$$

where

$$F_\pm(x, \cdot) = \int_0^\infty e^{\pm ix(\cosh v - 1)}g(v, \cdot)dv.$$

It suffices to show that $F_\pm(x, \cdot) = \frac{1}{x^{1/2}}W_\pm(x, \cdot) + O_A(x^{-A})$ with $W_\pm(x, \cdot)$ satisfying the required derivative bounds. For notational simplicity, we consider only the $+$ case, which we write as $F(x, \cdot)$ for $F_+(x, \cdot)$.

Write a smooth partition of unity of the form

$$1 = f_0(v/U) + \sum_V f_1(v/V) + f_2(v)   \text{for } v > 0,$$

with the following properties: f_0 has support on $[-1, 1]$, f_1 has support on $[1, 2]$, f_2 vanishes on $[0, 1]$, $U = x^{-1/2+\varepsilon}$, and V runs over $O(\log x)$ real numbers with $U \ll V \ll 1$. Repeated integration by parts shows that for all sufficiently large $A > 0$,

(10.10) $\int_1^\infty e^{ix(\cosh v - 1)}g(v, \cdot)f_2(v)\, dv \ll T^j x^{-j} \ll_A x^{-A},$

taking j large, and using $x \gg T^2$. Similarly, applying [BKY13, Lem. 8.1] with parameters $(X, Y, Q, R, U, \alpha, \beta)$ in our situation taking the values $(1, x, 1, xV, 1, V, 2V)$, we see that

$$\int_{-\infty}^\infty e^{ix(\cosh v - 1)}g(v, \cdot)f_1(v/V)\, dv \ll_A x^{-A}.$$

Hence

$$F(x, \cdot) = \int_0^2 e^{ix(\cosh v - 1)}g(v, \cdot)f_0\left(\frac{v}{U}\right)dv + O_A(x^{-A}).$$

Now let us develop $e^{ix(\cosh v - 1)}$ by first taking the Taylor expansion for $\cosh v - 1$ and then expanding it in the power series expansion for exp. We get that

$$e^{ix(\cosh v - 1)} = e^{ixv^2/2} \sum_{m=0}^{M} b_m \Big(x \sum_{n=0}^{N} a_n v^{2n+4} \Big)^m + O\Big(xv^{2N+6} + (xv^4)^{M+1} \Big).$$

Since $v \ll x^{-1/2+\varepsilon}$, we may take M, N large enough depending on ε, A so that

$$e^{ix(\cosh v - 1)} = e^{ixv^2/2} \sum_{m,n \geq 0} c_{m,n} (xv^2)^m v^{2n} + O_A(x^{-A})$$

for some new coefficients $c_{m,n}$ and where all but finitely many of the $c_{m,n}$ are zero.

Thus

$$F(x, \cdot) = \sum_{m \leq M, n \leq N} c_{m,n} \int_0^\infty (xv^2)^m v^{2n} e^{ixv^2/2} g(v, \cdot) f_0\Big(\frac{v}{U}\Big) dv + O_A(x^{-A}).$$

It transpires that g is nearly constant on the support of f_0. To see this, we note that

$$UT \ll x^{-\delta}$$

for some $\varepsilon > \delta > 0$, where ε is the ε appearing in $x \gg T^{2+\varepsilon}$, and we have chosen the ε in the definition of U small enough in comparison. Then, for any L, we have

$$g^{(L)}(\xi) v^L \ll (UT)^L \ll x^{-L\delta},$$

so that we may develop g in a Taylor series around 0 with finitely many terms, the number of which only depends on A, ε. Hence

$$F(x, \cdot) = \sum_{\ell \leq L, m \leq M, n \leq N} c_{\ell,m,n} g^{(\ell)}(0)$$
$$\times \int_0^\infty (xv^2)^m v^{2n+\ell} e^{ixv^2/2} f_0\Big(\frac{v}{U}\Big) dv + O_A(x^{-A})$$

for all sufficiently large L, M, N. Changing variables $v \to x^{-1/2} v$, we obtain
(10.11)
$$F(x, \cdot) = x^{-1/2} \sum_{\ell,m,n} c_{\ell,m,n} \frac{g^{(\ell)}(0)}{x^{\ell/2}} x^{-n} \int_0^\infty v^{2m+2n+\ell} e^{iv^2/2} f_0\Big(\frac{v}{x^\varepsilon}\Big) dv + O_A(x^{-A}).$$

Let us analyze the inner integral. We claim

$$\int_0^\infty v^N e^{iv^2/2} f_0\Big(\frac{v}{x^\varepsilon}\Big) dv = C(N) + O_{N,A}(x^{-A})$$

for some constant $C(N)$ independent of f_0 and x.

Proof of claim. For a smooth function f supported on $|v| \ll 1$, define

$$I(N, f, V) = \int_0^\infty v^N e^{iv^2/2} f(v/V) dv,$$

where $V \gg 1$ is large. Our first observation is that [BKY13, Lem. 8.1] shows that $I(N, f, V) \ll_{A,N,f} V^{-A}$ provided f is supported on $[1/2, 4]$, say. Our f_0 is not supported on this interval, but this argument shows $I(0, f_0, V) = I(0, 1, V) + O_A(V^{-A})$, where $I(0, 1, V) = e^{\pi i/4} \sqrt{\frac{\pi}{2}}$. Next, an integration by parts argument shows

$$I(N, f, V) = i\delta_{N=1} f(0) + i(N-1)I(N-2, f, V) + iV^{-1}I(N-1, f', V).$$

Here we interpret $I(M, f, V) = 0$ if $M < 0$. Since f_0' is dyadically-supported, this implies

$$I(N, f_0, x^\varepsilon) = i\delta_{N=1} + i(N-1)I(N-2, f_0, x^\varepsilon) + O_{N,A}(x^{-A}).$$

Repeating, we obtain the claim. □

Applying the claim to (10.11), we have

$$F(x, \cdot) = x^{-1/2} \sum_{\ell \leq L, m \leq M, n \leq N} c_{\ell,m,n} \frac{g^{(\ell)}(0)}{x^{\ell/2}} x^{-n} + O_A(x^{-A})$$

for some newly re-defined sequence of coefficients $c_{\ell,m,n}$, which completes the proof. □

10.3. *Properties of J^-.*

LEMMA 10.4. *We have*

$$(10.12) \qquad \frac{\partial^k}{\partial x^k} J^-(x, \cdot) \ll_{k,\varepsilon} x^{1-\varepsilon}(x^{-k} + x^k) T^{1+k+\varepsilon}.$$

Moreover, $J^-(x, \cdot)$ is 1-inert with respect to the variables $t_1, t_2, t_3, c, m_1, m_2, m_3$.

As in the J^+ case, this lemma is of interest to us when x is not too large. In the complementary case we have the following.

LEMMA 10.5. *Suppose for some $\varepsilon > 0$ that $1 \leq T^{1+\varepsilon} \ll x$. Then $J^-(x, \cdot) \ll_A x^{-A}$*

Proof of Lemma 10.4. As in the proof of Lemma 10.2, the property that J^- is 1-inert with respect to the other variables is easy to see, so we now focus on the bound (10.12). By [GR00, 8.486.10], we have

$$(10.13) \qquad \begin{aligned} J^-(x, \cdot) &= \frac{x}{i\pi^2} \int_{-\infty}^\infty (K_{1+2it}(x) - K_{1-2it}(x)) \sinh(\pi t) \exp(-(t/T)^2) \\ &\times \frac{(t^2 + \frac{1}{4})}{T^2} V_1(\cdot, t) V_2(\cdot, t) dt. \end{aligned}$$

From [GR00, 8.432.5], that is,

$$K_\nu(x) = \frac{\Gamma(\nu + \frac{1}{2})2^\nu}{x^\nu \Gamma(\frac{1}{2})} \int_0^\infty \frac{\cos(xt)}{(t^2+1)^{\frac{1}{2}+\nu}} dt, \qquad \mathrm{Re}(\nu) > 0,$$

one may readily deduce that

$$(10.14) \qquad K_{\varepsilon+2iy}(x) \ll_\varepsilon \frac{(1+|y|)^\varepsilon}{x^\varepsilon \cosh(\pi y)}$$

for $y \in \mathbb{R}$. For the part of the integral (10.13) with K_{1+2it} we move the contour to $\mathrm{Re}(1+2it) = \varepsilon > 0$, in all giving a contribution to $J^-(x)$ of size $\ll x^{1-\varepsilon}T^{1+\varepsilon}$. A similar bound works for the part of the integral with $K_{1-2it}(x)$, but by shifting the contour the other way. This gives the desired bound for $k = 0$.

Next we sketch how to treat $k \geq 1$. The bound on $K_{\varepsilon+2iy}$ given in (10.14) has the same essential features as (10.6). Moreover, the K-Bessel function satisfies (10.7). The same method used for J^+ now carries over to J^- without any significant changes. $\qquad\square$

Proof of Lemma 10.5. From [GR00, 8.432.4] one may derive

$$J^-(x,\cdot) = T^2 \int_{-\infty}^\infty \cos(x \sinh(v))g(v,\cdot)dv,$$

where g satisfies (10.9). (Here $g(v,\cdot)$ may differ slightly from that occurring in the proof of Lemma 10.3, but only by an absolute constant.)

As in the proof of Lemma 10.3, we can cut the integral at $|v| \leq 1$ again (with a smooth cutoff), since repeated integration by parts shows the complement is $O_A(x^{-A})$ for any $A > 0$. Therefore,

$$J^-(x,\cdot) = T^2 \int_{-\infty}^\infty \cos(x \sinh v)g_1(v,\cdot)dv + O_A(x^{-A}),$$

where $g_1^{(j)}(v,\cdot) \ll_A T^j(1+|v|)^{-A}$. We then change variables $v = \mathrm{arcsinh}(u)$, so $dv = (1+u^2)^{-1/2}du$, giving

$$(10.15) \qquad \begin{aligned} J^-(x,\cdot) &= T^2 \int_{-\infty}^\infty \cos(xu)g_0(u,\cdot)du + O_A(x^{-A}), \\ g_0(u,\cdot) &= g_1(\mathrm{arcsinh}(u),\cdot)(1+u^2)^{-1/2}. \end{aligned}$$

Since $\mathrm{arcsinh}(u)$ is smooth with bounded derivatives for $u \ll 1$, then $g_0(u,\cdot)$ is Schwartz-class and satisfies $g_0^{(j)}(u,\cdot) \ll_{j,A} T^j(1+|u|)^{-A}$. The integral in (10.15) is a cosine transform of g_0, and is hence $O(T^2(T/x)^j)$, for any $j \geq 0$, which is $O_A(x^{-A})$ for any $A > 0$, since $x \gg T^{1+\varepsilon}$ by assumption. $\qquad\square$

11. **Archimedean aspects, part 2**

The goal in this section is to understand the behavior of $K^{\pm}$ defined by (4.13).

We begin with some comments to help bridge the material in [PY19b, §10.4] to this work. In that article, the analog of K was defined but with $J^{\pm}(x, \cdot)$ replaced by $J_{\kappa-1}(x)$, the J-Bessel function, with κ fixed. Nevertheless, a great majority of the work done on K in [PY19b] carries over to this article, and the properties of $J^{\pm}$ developed in Section 10 will allow for this extension.

Throughout this section we assume that for some $0 < \eta \leq 1/13$,

$$(11.1) \qquad\qquad T \ll q^{\eta}.$$

The precise T-dependence is not important for the proof of Theorem 1.1.

11.1. *The properties of K.*

LEMMA 11.1 (Oscillatory Case). *Suppose that $|m_j| \asymp M_j$ for $j = 1, 2, 3$ and that $c \asymp C$. Suppose there exists $\delta > 0$ such that*

$$(11.2) \qquad\qquad \frac{\sqrt{N_1 N_2 N_3}}{C} \gg T^2 q^{\delta}.$$

Then
(11.3)
$$
K^{+}(m_1, m_2, m_3, c) = \frac{T^2 C^2 (N_1 N_2 N_3)^{1/2} e_c(-m_1 m_2 m_3)}{M_1 M_2 M_3} L(m_1, m_2, m_3, c)
$$
$$
+ O_{\delta, A}\Big(q^{-A} \prod_{j=1}^{3}(1 + m_j)^{-2}\Big),
$$

where L has the following properties. Firstly, L vanishes (meaning K^{+} is very small) unless

$$(11.4) \qquad\qquad M_j \asymp \frac{(N_1 N_2 N_3)^{1/2}}{N_j}, \quad j = 1, 2, 3,$$

and all the m_j have the same sign. Moreover, we have that

$$
(11.5) \quad L(m_1, m_2, m_3, c) = \int_{|\mathbf{u}| \ll q^{\varepsilon}} \int_{|y| \ll q^{\varepsilon}} F(\mathbf{u}; y)\Big(\frac{|m_1 m_2 m_3|}{c}\Big)^{iy}
$$
$$
\times \Big(\frac{M_1}{|m_1|}\Big)^{u_1} \Big(\frac{M_2}{|m_2|}\Big)^{u_2} \Big(\frac{M_3}{|m_3|}\Big)^{u_3} \Big(\frac{C}{c}\Big)^{u_4} d\mathbf{u} dy,
$$

where $F = F_{C, N_1, N_2, N_3, M_1, M_2, M_3}$ is entire in terms of $\mathbf{u}$ and satisfies

$$
F(\mathbf{u}; y) \ll_{\mathrm{Re}(\mathbf{u}), A} (1 + |\mathbf{u}|)^{-A}(1 + |y|)^{-A}
$$

for any $A > 0$.

Finally, $K^{-}(m_1, m_2, m_3, c) \ll q^{-1000}$.

Sketch of proof. The conditions in place in Lemma 11.1 ensure that the hypotheses of Lemma 10.3 are satisfied, which shows that $J^+(x)$ is essentially of the form $e^{\pm ix}/\sqrt{x}$, times a smooth function of x. This is the same shape of $J_{\kappa-1}(x)$ that was used in [PY19b, Lem. 11] and so the method used there carries over with minimal changes.

The final statement on the small size of K^- follows immediately from Lemma 10.5. $\qquad\square$

LEMMA 11.2 (Non-oscillatory case). *Suppose that* $m_j \asymp M_j$ *for* $j = 1, 2, 3$, $c \asymp C$, *and*

(11.6)
$$\frac{\sqrt{N_1 N_2 N_3}}{C} \ll T^2 q^\delta.$$

Then for both cases $K = K^\pm$, *we have*

(11.7)
$$K(m_1, m_2, m_3, c) = T N_1 N_2 N_3 \Big(\frac{\sqrt{N_1 N_2 N_3}}{C}\Big) e_c(-m_1 m_2 m_3) \int_{|\mathbf{u}| \ll T^2 q^{\delta+\varepsilon}} F(\mathbf{u})$$
$$\times \int_{|t| \ll q^\varepsilon + P} f(t) \Big(\frac{|m_1 m_2 m_3|}{c}\Big)^{it} \Big(\frac{M_1}{|m_1|}\Big)^{u_1} \Big(\frac{M_2}{|m_2|}\Big)^{u_2} \Big(\frac{M_3}{|m_3|}\Big)^{u_3} \Big(\frac{C}{c}\Big)^{u_4} dt\, d\mathbf{u}$$
$$+ O_{\varepsilon,A}\Big(q^{-A} \prod_{j=1}^3 (1 + m_j)^{-2}\Big),$$

where P *is defined by*

(11.8)
$$P = \frac{M_1 M_2 M_3}{C},$$

$f(t) \ll (1 + |t|)^{-1/2}$, *and* $F(\mathbf{u}) \ll_{A,\mathrm{Re}(\mathbf{u}),\varepsilon} q^\varepsilon \prod_{\ell=1}^4 (1 + \frac{|u_\ell|}{q^\varepsilon})^{-A}$ *for all* $A > 0$. *Moreover,* F *vanishes (meaning* K *is small) unless*

(11.9)
$$\frac{M_1 N_1}{C} \ll_\varepsilon q^\varepsilon, \qquad \frac{M_2 N_2}{C} \ll_\varepsilon q^\varepsilon, \qquad \frac{M_3 N_3}{C} \ll_\varepsilon q^\varepsilon.$$

If there exists $\varepsilon > 0$ *such that* $P \gg q^\varepsilon$, *then* f *may be chosen to have support on* $|t| \asymp P$.

Sketch of proof. In this case, $J^\pm$ satisfies the properties of Lemma 10.2 or 10.4 (depending on the choice of $\pm$). In turn, these are essentially the only properties that were used about $J_{\kappa-1}(x)$ in [PY19b, Lem. 12]. $\qquad\square$

LEMMA 11.3 (Other cases). *Suppose some* $m_j = 0$, *and let* K *denote either case of* $K^\pm$. *If* (11.2) *holds, then* K *is small. If* (11.6) *holds, then* K *is small unless* $|m_j| \ll_\varepsilon \frac{C}{N_j} q^\varepsilon$ *for* $j = 1, 2, 3$, *in which case*

(11.10)
$$K(m_1, m_2, m_3; c) \ll_\varepsilon T N_1 N_2 N_3 \Big(\frac{\sqrt{N_1 N_2 N_3}}{C}\Big) q^\varepsilon.$$

Proof. The fact that K is small if (11.2) holds follows from repeated integration by parts (see [BKY13, Lem. 8.1] for instance). If (11.6) holds, then another repeated integration by parts argument shows that the integral is small if there exists $\varepsilon > 0$ such that $|m_j| \gg \frac{C}{N_j}q^\varepsilon$ for some j. Finally, the bound (11.10) follows from trivially estimating the integral defining K, using (10.5) or (10.12). □

12. Completing the proof of Theorem 1.1

Here we finish the proof of the bounds $\mathcal{T}^\pm, \mathcal{T}_0^\pm \ll_\varepsilon T^B q^\varepsilon$ (for definitions, see (4.14) and (4.15)), which will complete the proof of Theorem 1.1.

We only deal with the case that $\epsilon_j = 1$ for all $j = 1, 2, 3$. The other cases are similar. Recall the definition of $\mathcal{T}^\pm$ from (4.14):

$$\mathcal{T}^\pm := \frac{1}{C\sqrt{N_1 N_2 N_3}} \sum_{\substack{m_1, m_2, m_3, r \geq 1 \\ (m_1, r) = 1}} G(m_1, m_2, m_3; qr) K^\pm(m_1, m_2, m_3, qr).$$

Using (5.3), we have

$$|\mathcal{T}^\pm| \ll \frac{1}{C^2 q \sqrt{N_1 N_2 N_3}}$$
$$\times \Big| \sum_{\substack{m_1, m_2, m_3, r \geq 1 \\ (m_1, r) = 1}} e_{qr}(m_1 m_2 m_3) K^\pm(m_1, m_2, m_3, qr) H_\chi(\pm m_1, m_2, m_3, r) \Big|.$$

Letting $N = N_1 N_2 N_3$, the behavior of K depends on whether or not

$$(12.1) \qquad\qquad \frac{\sqrt{N}}{C} \gg T^2 q^\varepsilon.$$

Oscillatory case. Suppose (12.1) holds for some $\varepsilon > 0$. By Lemma 11.1, only the case of K^+ is relevant, in which case we have (recalling (5.4))

$$(12.2) \qquad \begin{aligned} |\mathcal{T}^+| &\ll \frac{T^2}{qM} \Big| \int_{|\mathbf{u}| \ll q^\varepsilon} \int_{|y| \ll q^\varepsilon} F(\mathbf{u}; y) q^{-iy} M_1^{u_1} M_2^{u_2} M_3^{u_3} (C/q)^{u_4} \\ &\qquad \times Z(u_1 - iy, u_2 - iy, u_3 - iy, u_4 + iy) d\mathbf{u}\, dy \Big|, \end{aligned}$$

plus a small error term, where $M = M_1 M_2 M_3$. Here we initially take $\mathrm{Re}(u_j) = 1 + \varepsilon$ for all j. According to Lemma 8.1, write $Z = Z_0 + Z_1$. For Z_0, we keep the lines at $1 + \varepsilon$, while for Z_1 we move them to $1/2 + \varepsilon$. By the decay properties of F, the horizontal contour integrals arising from these contour shifts are small ($\ll q^{-100}$, say), and we will not mention them further. Thus we obtain

$$(12.3) \qquad \mathcal{T}^+ \ll \frac{q^\varepsilon T^2}{qM}\Big(\frac{MC}{q} + \frac{\sqrt{MC}}{\sqrt{q}} q^{3/2}\Big) \ll q^\varepsilon T^2\Big(\frac{C}{q^2} + \frac{\sqrt{C}}{N^{1/4}}\Big),$$

using that K^+ is very small unless $M \asymp \sqrt{N}$ in this oscillatory case. Since $CT^2 \ll N^{1/2} \ll (qT)^{3/2+\varepsilon}$ (from (12.1) and (4.8)), we have $\mathcal{T}^\pm \ll Tq^\varepsilon$ (using $T \ll q^\eta$ for some $\eta > 0$ small).

Non-oscillatory case. The method of estimation is similar in the case that $\frac{\sqrt{N}}{C} \ll T^2 q^\varepsilon$, but we use Lemma 11.2 in place of Lemma 11.1. From the terms with $m_j \asymp M_j$, we obtain that the contribution to $\mathcal{T}^\pm$ is

$$\ll \frac{NT}{C^3 q} \Big| \int_{|t| \ll q^\varepsilon + P} f(t) \int_{|\mathbf{u}| \ll T^2 q^\varepsilon} M_1^{u_1} M_2^{u_2} M_3^{u_3} \Big(\frac{C}{q}\Big)^{u_4}$$
$$\times \frac{F(\mathbf{u})}{q^{it}} Z(u_1 - it, u_2 - it, u_3 - it, u_4 + it) d\mathbf{u} dt \Big|,$$

where $P = M/C$. By the large sieve-like bound (8.1), we have that the contribution to the above from Z_1, say $\mathcal{T}_1^\pm$, satisfies the bound

$$\mathcal{T}_1^\pm \ll_\varepsilon \frac{NTq^\varepsilon}{C^3 q} \frac{\sqrt{MC}}{\sqrt{q}} q^{3/2} \Big(1 + \frac{\sqrt{M}}{\sqrt{C}}\Big) T^2.$$

In this case, $M \ll_\varepsilon \frac{C^3}{N} q^\varepsilon$, and so this bound becomes

(12.4) $$\mathcal{T}_1^\pm \ll_\varepsilon q^\varepsilon T^3 \Big(\frac{\sqrt{N}}{C} + 1\Big) \ll T^5 q^\varepsilon.$$

Next consider the contribution from Z_0, say $\mathcal{T}_{00}^\pm$. If $P \gg q^\varepsilon$ for some $\varepsilon > 0$, then we may assume f is supported on $|t| \asymp P$, and we shift the contours to the $(1/2 + \varepsilon)$-line. No poles are crossed during this procedure since they occur at height t, and the horizontal integrals arising from this contour shift are negligible since F is small at this height. By the final sentence of Lemma 8.1, the bound we obtain on $\mathcal{T}_{00}^\pm$ is no worse than the bound on $\mathcal{T}_1^\pm$ given in (12.4).

Finally, consider the case $P \ll_\varepsilon q^\varepsilon$, that is, $M \ll_\varepsilon Cq^\varepsilon$. Here we keep the contours at the $(1 + \varepsilon)$-line, giving

$$\mathcal{T}_{00}^\pm \ll_\varepsilon \frac{NT}{C^3 q} \frac{MC}{q} T^2 q^\varepsilon \ll \frac{NT^3}{Cq^2} q^\varepsilon = T^3 \frac{\sqrt{N}}{C} \frac{\sqrt{N}}{q^2} q^\varepsilon \ll T^{\frac{13}{2}} q^{-\frac{1}{2}+\varepsilon},$$

using (4.8), which is $\ll_\varepsilon q^\varepsilon$ taking $\eta \leq 1/13$ in (11.1).

The cases with some $m_j = 0$. We will estimate $\mathcal{T}_0^\pm$ by trivial bounds. By Lemma 11.3, $K(m_1, m_2, m_3, c)$ is very small in this case, unless we are in the non-oscillatory situation (11.6).

Using Lemma 11.3, we deduce

(12.5) $$\mathcal{T}_0^\pm \ll_\varepsilon \frac{TN}{C^2} q^\varepsilon \sum_{r \asymp C/q} \sum_{\substack{m_1 m_2 m_3 = 0 \\ |m_j| \ll_\varepsilon M_j}} |G(m_1, m_2, m_3; qr)|, \qquad M_j := \frac{C}{N_j} q^\varepsilon,$$

plus a small error term. Recall the bound (6.9), and that $G(m_1, m_2, m_3, qr) = 0$ if $(m_1, r) \neq 1$.

First consider the terms with $m_3 = 0$ and $m_1, m_2 \neq 0$. Their contribution to (12.5) is

$$(12.6) \qquad \ll_\varepsilon \frac{TN}{C^2} \frac{C}{q} M_1 M_2 \frac{q^\varepsilon}{Cq} \ll T \frac{N_3}{q^2} q^\varepsilon \ll T^3 q^\varepsilon,$$

using (4.8). The case with $m_2 = 0$ and $m_1, m_3 \neq 0$ is essentially identical to the previous case, but the case with $m_1 = 0$ and $m_2, m_3 \neq 0$ is slightly different because of the condition $(m_1, r) = 1$. The r-sum collapses to $r = 1$, and this sum is even smaller than that appearing in the previous cases (essentially, the factor $\frac{C}{q}$ may be improved to 1).

Next consider the terms with two $m_j = 0$, the hardest one being $m_2 = m_3 = 0$. Compared to (12.6), the difference is that the factor M_2 is replaced by q, leading to the bound

$$\ll_\varepsilon \frac{TN}{C^2} \frac{C}{q} M_1 q \frac{q^\varepsilon}{Cq} \ll T \frac{N_2 N_3}{qC} q^\varepsilon \ll T \frac{\sqrt{N}}{C} \frac{\sqrt{N_2 N_3}}{q} q^\varepsilon.$$

Using $\frac{\sqrt{N}}{C} \ll_\varepsilon T^2 q^\varepsilon$ and $N_2 N_3 \ll_\varepsilon (qT)^{2+\varepsilon}$ (recall (4.8)) shows this is $\ll_\varepsilon T^4 q^\varepsilon$. If m_1 is one of the two m_j's equal to zero, then the numerology changes enough to be worthy of mention. (We no longer have $N_1 N_3 \ll_\varepsilon (Tq)^{2+\varepsilon}$, but on the other hand the r-sum collapses, so we may assume $C \asymp q$ since $c = qr \asymp C$.) Say $m_1 = m_3 = 0$ and $m_2 \neq 0$. Then the contribution of these terms to $\mathcal{T}_0^\pm$ is

$$\ll_\varepsilon \frac{TN}{C^2} q M_2 \frac{q^\varepsilon}{Cq} \ll \frac{TN}{C^3} M_2 q^\varepsilon \ll \frac{TN_1 N_3}{C^2} q^\varepsilon \ll T^5 q^\varepsilon,$$

where we used $\frac{C}{q} \asymp 1$, $M_2 \ll \frac{C}{N_2} q^\varepsilon$, and $N_1 N_3 \leq N \ll C^2 T^4 q^\varepsilon$.

Finally, the terms of $\mathcal{T}_0^\pm$ with $m_1 = m_2 = m_3 = 0$ (hence $r = 1$, $C \asymp q$) are bounded by

$$\ll_\varepsilon \frac{TN}{C^2} \frac{q}{C} q^\varepsilon \ll T^5 q^\varepsilon.$$

This completes the proof of Theorem 1.1.

13. Sketch of proof of Theorems 1.2 and 1.5

In this section, we outline what changes are needed to prove Theorem 1.2. The problem is arithmetically identical to the proof of Theorem 1.1, but the Archimedean aspects are different. Recall we have assumed that $T \gg q^\eta$ for some small but fixed $\eta > 0$.

The first change is that instead of using $h_0(t)$ defined by (4.1), we take

$$h_0(t) = \frac{1}{\cosh\left(\frac{t-T}{\Delta}\right)} + \frac{1}{\cosh\left(\frac{t+T}{\Delta}\right)},$$

as in [You17, §4], where $\Delta = T^\varepsilon$ for some $\varepsilon > 0$. A more precise version of Lemma 10.1 is developed in [You17, §5], showing that $V_j(y, t)$ has an asymptotic expansion with leading term of the form $W_j(\frac{y}{T^j})$, where W_1 and

W_2 are fixed smooth weight functions, satisfying $x^k W_j^{(k)}(x) \ll_A (1 + |x|)^{-A}$ for all $A > 0$. The analogs of the estimates for $J^{\pm}$ appear as [You17, Lem. 7.1, 7.2], while the crucial integral representations of $K(m_1, m_2, m_3, c)$ are treated in [You17, Lem. 8.1] in place of those covered in Section 11. Note that in [You17, (8.5)], the contours were set at $\mathrm{Re}(\mathbf{y}) = \mathrm{Re}(u) = 0$. To accommodate more general choices of contour, the formula [You17, (8.4)] should be updated to state

$$K^+(m_1, m_2, m_3, c) = \frac{C^{3/2} \Delta T (N_1 N_2 N_3)^{1/2} e_c(-m_1 m_2 m_3)}{(M_1 M_2 M_3)^{1/2}} L(m_1, m_2, m_3, c),$$

plus a small error term, where

$$L(m_1, m_2, m_3, c) = \frac{1}{V} \int_{|\mathbf{u}| \ll (qT)^\varepsilon} \int_{|y| \ll U} F(\mathbf{u}; y) \left(\frac{|m_1 m_2 m_3|}{c} \right)^{iy}$$
$$\times \left(\frac{M_1}{|m_1|} \right)^{u_1} \left(\frac{M_2}{|m_2|} \right)^{u_2} \left(\frac{M_3}{|m_3|} \right)^{u_3} \left(\frac{C}{c} \right)^{u_4} d\mathbf{u}\, dy,$$

where $V = T$ and

$$U = \frac{T^2 C}{(N_1 N_2 N_3)^{1/2}}.$$

Moreover, L vanishes (i.e., K^+ is very small) unless

$$C \ll_\varepsilon \frac{(N_1 N_2 N_3)^{1/2}}{\Delta^{1-\varepsilon} T} \qquad \text{and} \qquad M_j \asymp \frac{(N_1 N_2 N_3)^{1/2}}{N_j}, j = 1, 2, 3.$$

The formula for K^- can be adapted in a similar way, but we leave out the details for brevity.

Now if we follow along the details of the *Oscillatory case* from Section 12, we obtain that the contribution to $\mathcal{T}^\pm$ from these terms is (in place of (12.2))

$$|\mathcal{T}^\pm| \ll \frac{\Delta T}{q C^{1/2} M^{1/2} V} \Big| \int_{|\mathbf{u}| \ll q^\varepsilon} \int_{|y| \ll U} F(\mathbf{u}; y) q^{-iy} M_1^{u_1} M_2^{u_2} M_3^{u_3} (C/q)^{u_4}$$
$$\times Z(u_1 - iy, u_2 - iy, u_3 - iy, u_4 + iy) d\mathbf{u}\, dy \Big|,$$

plus a small error term. We decompose Z as $Z_0 + Z_1$, and for Z_1, we shift the contour to the $(1/2 + \varepsilon)$-lines, giving that its contribution to $\mathcal{T}^\pm$ is

$$\ll_\varepsilon \frac{1}{q} \frac{\Delta T}{C^{1/2} M^{1/2}} \frac{U}{V} \frac{\sqrt{MC}}{\sqrt{q}} q^{3/2} T^\varepsilon.$$

Using $\frac{U}{V} \ll_\varepsilon \Delta^{-1+\varepsilon}$ shows this term is $\ll_\varepsilon T^{1+\varepsilon}$, which is the bound required for Theorem 1.2. Next we turn to Z_0. For this term, it is helpful to point out that in fact $F(\mathbf{u}; y)$ is very small unless $|y| \asymp U$, which was a property that was not stated in [You17, Lem. 8.1], but was developed in the proof (see [You17, p.1569]). This shows that if $U \gg T^\varepsilon$ for some $\varepsilon > 0$, then in the estimation of Z_0 we can shift the contours to the $(1/2 + \varepsilon)$-lines without

crossing poles. The bound obtained on Z_0 is no larger than the one obtained on Z_1. If $U \ll_\varepsilon T^\varepsilon$, then we keep the contours at the $(1 + \varepsilon)$-lines, giving that their contribution to $\mathcal{T}^\pm$ is

$$\ll_\varepsilon \frac{\Delta T}{qC^{1/2}M^{1/2}V} \frac{MC}{q} T^\varepsilon \ll q^{-1/2}\Delta^{1/2}T^{1+\varepsilon},$$

which is stronger than the bound obtained on Z_1.

The *Non-oscillatory case* is similar, and we omit the details for brevity.

Finally, we need to consider the terms where some $m_j = 0$. These cases were overlooked in [You17], so we take this opportunity to correct this omission. The first claim is that $K^+(m_1, m_2, m_3, c)$ is very small if some $m_j = 0$. This follows from the fact that $B^+(x)$ (the analog of $J^+(x, \cdot)$) is very small unless $x \gg_\varepsilon \Delta T^{1-\varepsilon}$, in which case it has an asymptotic expansion of the form $\frac{\Delta T}{\sqrt{x}} \cos(x + \phi(x, T))$, where $\phi(x, T) = -2T^2/x + \cdots$. Then repeated integration by parts in the t_j variable (where $m_j = 0$) shows that K^+ is small. Therefore, it suffices to consider K^-. We claim that if some $m_j = 0$, then

(13.1) $$K^-(m_1, m_2, m_3, c) \ll_\varepsilon \Delta N T^\varepsilon.$$

The trivial bound arising from [You17, Lem. 7.2] would give a bound of the form NT, so (13.1) saves a factor of T/Δ over this. We now prove the claim. According to [You17, (7.3)], we have

$$B^-(x) = \Delta T \int_{|v| \le \Delta^{-1+\varepsilon}} \cos(x \sinh v)e^{2ivT}g(\Delta v)dv + O_A(T^{-A}),$$

where $g^{(j)}(x) \ll_A (1 + |x|)^{-A}$ for all $A > 0$. Moreover, $B^-(x)$ is very small unless $x \asymp T$. Here $B^-(x)$ is the analog of $J^-(x, \cdot)$. To fix the notation, say $m_3 = 0$. (The cases with $m_1 = 0$ or $m_2 = 0$ are identical.) Then the t_3-integral inside the definition of K^- takes the form

$$\int_{-\infty}^\infty w(t_3, \cdot) \cos\left(\frac{4\pi\sqrt{t_1 t_2 t_3}}{c} \sinh v\right)dt_3,$$

where $w(t_3, \cdot)$ is supported on $t_3 \asymp N_3$, and satisfies $t_3^j \frac{d^j}{dt_3^j}w(t_3, \cdot) \ll 1$. Repeated integration by parts (see [BKY13, Lem. 8.1]) therefore shows that $K^-(m_1, m_2, m_3, c)$ is very small unless

$$\frac{\sqrt{N}}{C}|v| \ll_\varepsilon T^\varepsilon.$$

On the other hand, we also know K^- is very small unless $x \asymp \frac{\sqrt{N}}{C} \asymp T$, so inside the definition of K^- we may further restrict v by $|v| \ll_\varepsilon T^{-1+\varepsilon}$. The trivial bound on K^- now leads to (13.1). An integration by parts argument in the t_1, t_2 variables shows that $K^-(m_1, m_2, 0, c)$ is very small unless $|m_j| \ll_\varepsilon \frac{C}{N_j}T^\varepsilon$ for $j = 1, 2$.

At this point, we carry through the same argument used in Section 12, using (6.9) as before, but using (13.1) in place of Lemma 11.3. As a representative sample, consider the contribution from $m_3 = 0$, $m_1, m_2 \neq 0$. These terms give

$$\ll_\varepsilon \frac{1}{C\sqrt{N}} \frac{\Delta N}{Cq} T^\varepsilon \sum_{r \asymp \frac{C}{q}} \sum_{1 \leq |m_1| \ll_\varepsilon \frac{C}{N_1} T^\varepsilon} \sum_{1 \leq |m_2| \ll_\varepsilon \frac{C}{N_2} T^\varepsilon} (m_2, q)(m_3, q) \ll \frac{\Delta N_3}{q^2 T} T^\varepsilon,$$

using $C \asymp \frac{\sqrt{N}}{T}$. Since $N_3 \ll_\varepsilon (qT)^{2+\varepsilon}$, this is $\ll_\varepsilon \Delta T^{1+\varepsilon}$, which is the bound required for Theorem 1.2. Similar arguments may be used to treat the other terms with $m_1 m_2 m_3 = 0$, and we leave the details to the diligent reader.

The proof of Theorem 1.2 is now complete. $\qquad\square$

Finally, we discuss the proof of Theorem 1.5. The framework of [You17] placed both the Maass forms and holomorphic forms on an equal footing, and so the proof of the hybrid bound (1.8) is now essentially identical to that of Theorem 1.2. In order to derive the bound (1.7), one may adapt the material from Section 10. It is not difficult to prove an analogous version of Lemma 10.1. (The use of Stirling's formula is slightly different.) The use of the Bruggeman-Kuznetsov formula will then be replaced by the Petersson formula and Poisson summation over κ (see [Iwa97, pp. 85, 86]). One can then derive properties of the resulting weight functions that are analogous to those of $J^\pm$ presented in Sections 10.2 and 10.3. The properties of $K^\pm$ derived in Section 11 then carry over with minimal changes, and the final steps of Section 12 then proceed in the same fashion as in the proof of Theorem 1.1.

References

[AL70]   A. O. L. ATKIN and J. LEHNER, Hecke operators on $\Gamma_0(m)$, *Math. Ann.* **185** (1970), 134–160. MR 0268123. Zbl 0177.34901. https://doi.org/10.1007/BF01359701.

[AL78]   A. O. L. ATKIN and W. C. W. LI, Twists of newforms and pseudo-eigenvalues of W-operators, *Invent. Math.* **48** no. 3 (1978), 221–243. MR 0508986. Zbl 0369.10016. https://doi.org/10.1007/BF01390245.

[BLT64] M. B. BARBAN, Y. V. LINNIK, and N. G. TSHUDAKOV, On prime numbers in an arithmetic progression with a prime-power difference, *Acta Arith.* **9** (1964), 375–390. MR 0171766. Zbl 0127.26901. https://doi.org/10.4064/aa-9-4-375-390.

[BH08]   V. BLOMER and G. HARCOS, Hybrid bounds for twisted L-functions MR 2431250, *J. Reine Angew. Math.* **621** (2008), 53–79. MR 2431250. Zbl 1193.11044. https://doi.org/10.1515/crelle.2008.058.

[BH14]   V. BLOMER and G. HARCOS, Addendum: Hybrid bounds for twisted L-functions [MR 2431250], *J. Reine Angew. Math.* **694** (2014), 241–244. MR 3259045. Zbl 1296.11047. https://doi.org/10.1515/crelle-2012-0091.

[BHM07]   V. Blomer, G. Harcos, and P. Michel, A Burgess-like subconvex
          bound for twisted L-functions, *Forum Math.* **19** no. 1 (2007), 61–105, Ap-
          pendix 2 by Z. Mao. MR 2296066. Zbl 1168.11014. https://doi.org/10.
          1515/FORUM.2007.003.

[BKY13]   V. Blomer, R. Khan, and M. Young, Distribution of mass
          of holomorphic cusp forms, *Duke Math. J.* **162** no. 14 (2013),
          2609–2644. MR 3127809. Zbl 1312.11028. https://doi.org/10.1215/
          00127094-2380967.

[BM15]   V. Blomer and D. Milićević, *p*-adic analytic twists and strong sub-
          convexity, *Ann. Sci. Éc. Norm. Supér.* (4) **48** no. 3 (2015), 561–605.
          MR 3377053. Zbl 1401.11095. https://doi.org/10.24033/asens.2252.

[Bur63]   D. A. Burgess, On character sums and L-series. II, *Proc. London Math.
          Soc.* (3) **13** (1963), 524–536. MR 0148626. Zbl 0123.04404. https://doi.
          org/10.1112/plms/s3-13.1.524.

[CI00]   J. B. Conrey and H. Iwaniec, The cubic moment of central values of
          automorphic L-functions, *Ann. of Math.* (2) **151** no. 3 (2000), 1175–1216.
          MR 1779567. Zbl 0973.11056. https://doi.org/10.2307/121132.

[Del77]   P. Deligne, *Cohomologie Étale*, *Lecture Notes in Math.* **569**, Springer-
          Verlag, Berlin, 1977, Séminaire de géométrie algébrique du Bois-
          Marie SGA $4\frac{1}{2}$. MR 0463174. Zbl 0349.14008. https://doi.org/10.1007/
          BFb0091516.

[Del80]   P. Deligne, La conjecture de Weil. II, *Inst. Hautes Études Sci. Publ.
          Math.* no. 52 (1980), 137–252. MR 0601520. Zbl 0456.14014. Available at
          http://www.numdam.org/item?id=PMIHES_1980__52__137_0.

[FKM21]   É. Fouvry, E. Kowalski, and P. Michel, On the conductor of coho-
          mological transforms, *Ann. Fac. Sci. Toulouse Math.* (6), 2021, to appear.
          arXiv 1310.3603.

[FKMS19] É. Fouvry, E. Kowalski, P. Michel, and W. Sawin, Lectures on
          applied ℓ-adic cohomology, in *Analytic Methods in Arithmetic Geom-
          etry*, *Contemp. Math.* **740**, Amer. Math. Soc., Providence, RI, 2019,
          pp. 113–195. MR 4033731. Zbl 07217729. https://doi.org/10.1090/conm/
          740/14903.

[Gel75]   S. S. Gelbart, *Automorphic Forms on Adele Groups*, *Ann. of Math.
          Stud.* **83**, Princeton Univ. Press, Princeton, N.J.; Univ. of Tokyo Press,
          Tokyo, 1975. MR 0379375. Zbl 0329.10018. https://doi.org/10.1515/
          9781400881611.

[GR00]   I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals*, *Series*, *and
          Products*, sixth ed., Academic Press, Inc., San Diego, CA, 2000, (translated
          from the Russian; translation edited and with a preface by Alan Jeffrey
          and Daniel Zwillinger). MR 1773820. Zbl 0981.65001.

[Guo96]   J. Guo, On the positivity of the central critical values of automorphic L-
          functions for GL(2), *Duke Math. J.* **83** no. 1 (1996), 157–190. MR 1388847.
          Zbl 0861.11032. https://doi.org/10.1215/S0012-7094-96-08307-6.

[HB78]   D. R. HEATH-BROWN, Hybrid bounds for Dirichlet L-functions, *Invent. Math.* **47** no. 2 (1978), 149–170. MR 0485727. Zbl 0362.10035. https://doi.org/10.1007/BF01578069.

[HL94]   J. HOFFSTEIN and P. LOCKHART, Coefficients of Maass forms and the Siegel zero, with an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman, *Ann. of Math.* (2) **140** no. 1 (1994), 161–181. MR 1289494. Zbl 0814.11032. https://doi.org/10.2307/2118543.

[Ivi01]   A. IVIĆ, On sums of Hecke series in short intervals, *J. Théor. Nombres Bordeaux* **13** no. 2 (2001), 453–468. MR 1879668. Zbl 0994.11020. https://doi.org/10.5802/jtnb.333.

[Iwa90]   H. IWANIEC, Small eigenvalues of Laplacian for $\Gamma_0(N)$, *Acta Arith.* **56** no. 1 (1990), 65–82. MR 1067982. Zbl 0702.11034. https://doi.org/10.4064/aa-56-1-65-82.

[Iwa97]   H. IWANIEC, *Topics in Classical Automorphic Forms*, Grad. Stud. Math. **17**, Amer. Math. Soc., Providence, RI, 1997. MR 1474964. Zbl 0905.11023. https://doi.org/10.1090/gsm/017.

[IK04]   H. IWANIEC and E. KOWALSKI, *Analytic Number Theory*, Amer. Math. Soc. Colloq. Publ. **53**, Amer. Math. Soc., Providence, RI, 2004. MR 2061214. Zbl 1059.11001. https://doi.org/10.1090/coll/053.

[JL70]   H. JACQUET and R. P. LANGLANDS, *Automorphic Forms on* GL(2), Lecture Notes in Math. **114**, Springer-Verlag, New York, 1970. MR 0401654. Zbl 0236.12010. https://doi.org/10.1007/BFb0058988.

[Kat88]   N. M. KATZ, *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, Ann. of Math. Stud. **116**, Princeton Univ. Press, Princeton, NJ, 1988. MR 0955052. Zbl 0675.14004. https://doi.org/10.1515/9781400882120.

[Kat01]   N. M. KATZ, Sums of Betti numbers in arbitrary characteristic, *Finite Fields Appl.* **7** no. 1 (2001), 29–44, dedicated to Professor Chao Ko on the occasion of his 90th birthday. MR 1803934. Zbl 1068.14501. https://doi.org/10.1006/ffta.2000.0303.

[KPY19]   E. M. KIRAL, I. PETROW, and M. P. YOUNG, Oscillatory integrals with uniformity in parameters, *J. Théor. Nombres Bordeaux* **31** no. 1 (2019), 145–159. MR 3994723. https://doi.org/10.5802/jtnb.1072.

[KL13]   A. KNIGHTLY and C. LI, Kuznetsov's trace formula and the Hecke eigenvalues of Maass forms, *Mem. Amer. Math. Soc.* **224** no. 1055 (2013), vi+132. MR 3099744. Zbl 1314.11038. https://doi.org/10.1090/S0065-9266-2012-00673-3.

[KMS17]   E. KOWALSKI, P. MICHEL, and W. SAWIN, Bilinear forms with Kloosterman sums and applications, *Ann. of Math.* (2) **186** no. 2 (2017), 413–500. MR 3702671. Zbl 06774111. https://doi.org/10.4007/annals.2017.186.2.2.

[Lau81]   G. LAUMON, Semi-continuité du conducteur de Swan (d'après P. Deligne), (French), in *The Euler-Poincaré Characteristic*, Astérisque **83**, Soc. Math. France, Paris, 1981, pp. 173–219. MR 0629128. Zbl 0504.14013. Available at http://www.numdam.org/item/AST_1981__82-83__173_0/.

[Li80]   W.-C. W. LI, On the representations of GL(2). I. ε-factors and n-closeness, *J. Reine Angew. Math.* **313** (1980), 27–42. MR 0552460. Zbl 0411.20023. https://doi.org/10.1515/crll.1980.313.27.

[Mil16]   D. MILIĆEVIĆ, Sub-Weyl subconvexity for Dirichlet L-functions to prime power moduli, *Compos. Math.* **152** no. 4 (2016), 825–875. MR 3484115. https://doi.org/10.1112/S0010437X15007381.

[Pet18]   I. PETROW, Bounds for traces of Hecke operators and applications to modular and elliptic curves over a finite field, *Algebra Number Theory* **12** no. 10 (2018), 2471–2498. MR 3911137. Zbl 07026824. https://doi.org/10.2140/ant.2018.12.2471.

[PY19a]   I. PETROW and M. P. YOUNG, The fourth moment of Dirichlet L-functions along a coset and the Weyl bound, 2019. arXiv 1908.10346.

[PY19b]   I. PETROW and M. P. YOUNG, A generalized cubic moment and the Petersson formula for newforms, *Math. Ann.* **373** no. 1-2 (2019), 287–353. MR 3968874. Zbl 07051745. https://doi.org/10.1007/s00208-018-1745-1.

[Pet15]   I. N. PETROW, A twisted Motohashi formula and Weyl-subconvexity for L-functions of weight two cusp forms, *Math. Ann.* **363** no. 1-2 (2015), 175–216. MR 3394377. Zbl 1331.11035. https://doi.org/10.1007/s00208-014-1166-8.

[Wal85]   J.-L. WALDSPURGER, Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie, *Compositio Math.* **54** no. 2 (1985), 173–242. MR 0783511. Zbl 0567.10021. Available at http://www.numdam.org/item?id=CM_1985__54_2_173_0.

[Wu14]   H. WU, Burgess-like subconvex bounds for $GL_2 \times GL_1$, *Geom. Funct. Anal.* **24** no. 3 (2014), 968–1036. MR 3213837. Zbl 1376.11037. https://doi.org/10.1007/s00039-014-0277-4.

[Wu19]   H. WU, Burgess-like subconvexity for GL_1, *Compos. Math.* **155** no. 8 (2019), 1457–1499. MR 3977317. Zbl 07077744. https://doi.org/10.1112/s0010437x19007309.

[You17]   M. P. YOUNG, Weyl-type hybrid subconvexity bounds for twisted L-functions and Heegner points on shrinking sets, *J. Eur. Math. Soc.* (*JEMS*) **19** no. 5 (2017), 1545–1576. MR 3635360. Zbl 1430.11067. https://doi.org/10.4171/JEMS/699.

[You19]   M. P. YOUNG, Explicit calculations with Eisenstein series, *J. Number Theory* **199** (2019), 1–48. MR 3926186. Zbl 07030973. https://doi.org/10.1016/j.jnt.2018.11.007.

ETH ZÜRICH, ZÜRICH, SWITZERLAND
E-mail: ian.petrow@math.ethz.ch

TEXAS A&M UNIVERSITY, COLLEGE STATION, TX, USA
E-mail: myoung@math.tamu.edu