

# THE FOURTH MOMENT OF DIRICHLET $L$ -FUNCTIONS ALONG A COSET AND THE WEYL BOUND

IAN PETROW AND MATTHEW P. YOUNG

**ABSTRACT.** We prove a Lindelöf-on-average upper bound for the fourth moment of Dirichlet  $L$ -functions of conductor  $q$  along a coset of the subgroup of characters modulo  $d$  when  $q^*|d$ , where  $q^*$  is the least positive integer such that  $q^2|(q^*)^3$ . As a consequence, we finish the previous work of the authors and establish a Weyl-strength subconvex bound for all Dirichlet  $L$ -functions with no restrictions on the conductor.

## 1. INTRODUCTION

**1.1. The Weyl bound and cubic moments.** This paper continues the previous work of the authors [PY2] on the Weyl bound for Dirichlet  $L$ -functions of cube-free conductor. In the present paper, we remove the cube-free hypothesis and establish the following theorem without any restrictions on the conductor of  $\chi$ .

**Theorem 1.1.** *For any primitive Dirichlet character  $\chi$  modulo  $q$  and  $\varepsilon > 0$ , we have*

$$(1.1) \quad L(1/2 + it, \chi) \ll_{\varepsilon} (q(1 + |t|))^{1/6+\varepsilon}.$$

In another language, for any Hecke character  $\chi$  over  $\mathbb{Q}$ , we have  $L(1/2, \chi) \ll_{\varepsilon} C(\chi)^{1/6+\varepsilon}$  where  $C(\chi)$  is the analytic conductor of  $\chi$ .

As in our previous work [PY2] and that of Conrey and Iwaniec [CI], Theorem 1.1 is based on Lindelöf-on-average upper bounds for two closely-related cubic moments, see Theorems 1.2 and 1.3 below. Let  $\mathcal{H}_{it}(m, \psi)$  denote the set of Hecke-Maass newforms of conductor  $m$ , central character  $\psi$ , and spectral parameter  $it$ . A key new idea in [PY2] was the shape of the family of automorphic forms into which we embed  $\chi$ , motivated by the following fact: If  $\chi$  is a primitive character modulo  $q$ ,  $m | q$  and  $\pi \in \mathcal{H}_{it}(m, \bar{\chi}^2)$ , then  $\pi \otimes \chi \in \mathcal{H}_{it}(q^2, 1)$ , see [JL, Prop. 3.8(iii)] or [AtLi, Thm. 3.1(ii)].

**Theorem 1.2.** *There exists a  $B > 2$  such that for all primitive  $\chi$  modulo  $q$  not quadratic and  $\varepsilon > 0$  we have*

$$(1.2) \quad \sum_{|t_j| \leq T} \sum_{m|q} \sum_{\pi \in \mathcal{H}_{it_j}(m, \bar{\chi}^2)} L(1/2, \pi \otimes \chi)^3 + \int_{-T}^T |L(1/2 + it, \chi)|^6 dt \ll_{\varepsilon} T^B q^{1+\varepsilon}.$$

**Theorem 1.3.** *For all primitive  $\chi$  modulo  $q$ ,  $\delta, \varepsilon > 0$ , and  $T \gg q^{\delta}$  we have*

$$(1.3) \quad \sum_{T \leq t_j < T+1} \sum_{m|q} \sum_{\pi \in \mathcal{H}_{it_j}(m, \bar{\chi}^2)} L(1/2, \pi \otimes \chi)^3 + \int_T^{T+1} |L(1/2 + it, \chi)|^6 dt \ll_{\delta, \varepsilon} T^{1+\varepsilon} q^{1+\varepsilon}.$$

---

2010 *Mathematics Subject Classification.* 11M06 (primary) 11F11 11F12 11F66.

The first author was supported by Swiss National Science Foundation grant PZ00P2\_168164. This material is based upon work supported by the National Science Foundation under agreement No. DMS-170222 (M.Y.). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

These two theorems, with the additional hypothesis that  $q$  is cube-free, appeared as Theorems 1.1 and 1.2 of [PY2]. That  $\pi \otimes \chi$  has trivial central character is crucial because we may then rely on deep results of Guo [G], which guarantee that  $L(1/2, \pi \otimes \chi) \geq 0$ . We then may conclude Theorem 1.1 by a standard positivity argument.

The reader may wonder why the cube-free hypothesis arose in our previous work and how we are able to remove it in this paper. In order to answer these questions, we briefly recall the proof of Theorems 1.1 and 1.2 of [PY2].

The proof of Theorem 1.2 begins with several standard steps to estimate (1.2). We apply an approximate functional equation to expand  $L(1/2, \pi \otimes \chi)$  as a finite sum, and apply the Bruggeman-Kuznetsov formula and Poisson summation. The result is a sum of complete character sums. The archimedean integral can be treated by the method of stationary phase, and the non-archimedean sum by an explicit elementary calculation. By Mellin inversion, the result of these steps is that the cubic moment (1.2) is transformed to a main term plus a reciprocal ‘‘dual moment’’ of the rough shape

$$(1.4) \quad \sum_{\psi \pmod{q}} |L(1/2, \psi)|^4 g(\chi, \psi),$$

where  $g(\chi, \psi)$  is defined by

$$(1.5) \quad g(\chi, \psi) = \sum_{u, t \pmod{q}} \chi(t) \bar{\chi}(t+1) \bar{\chi}(u) \chi(u+1) \psi(ut-1).$$

The existence of such a formula was first noticed in the case that  $\chi$  is quadratic by the first author in [P1]. There have been several other examples of such reciprocal dual moments that have underpinned many other results in the literature. For instance, Motohashi [Mo] proved a formula of the rough shape

$$\int w(t) |\zeta(1/2 + it)|^4 dt \leftrightarrow \sum_{t_j} \sum_{\pi \in \mathcal{H}_{it_j}(1,1)} \check{w}(t_j) L(1/2, \pi)^3,$$

with an explicit transform  $\check{w}$  of the test function  $w$ . See also [MV] for an elegant geometric proof of a special case and [Nel] for the generalization of their work to a wide class of test functions. In [Y1] the second author derived a similar duality in  $q$ -aspect:

$$\sum_{\chi \pmod{p}} |L(1/2, \chi)|^4 \leftrightarrow \sum_{t_j} \sum_{\pi \in \mathcal{H}_{it_j}(1,1)} \lambda_\pi(p) L(1/2, \pi)^3$$

We also mention the more recent papers [AK, BK, Fr, Z] giving additional reciprocity results for moments of  $L$ -functions.

To prove the estimates in Theorems 1.2 and 1.3, it suffices to show for all  $\varepsilon > 0$  that

$$(1.6) \quad \sum_{\psi \pmod{q}} |L(1/2, \psi)|^4 g(\chi, \psi) \ll_\varepsilon q^{2+\varepsilon}.$$

The sum  $g(\chi, \psi)$  is multiplicative, so it suffices to consider  $g(\chi, \psi)$  for  $q$  a prime power. If  $q = p$  is prime then the bound  $g(\chi, \psi) \ll p$  follows from the theory of  $\ell$ -adic sheaves and trace functions, and in particular the Riemann hypothesis of Deligne. If  $q = p^2$  then  $g(\chi, \psi) \ll p^2$  by an elementary calculation (see [PY2, §9.2]). In these cases, we have for all  $\varepsilon > 0$

$$\sum_{\psi \pmod{q}} |L(1/2, \psi)|^4 g(\chi, \psi) \ll_\varepsilon q^{1+\varepsilon} \sum_{\psi \pmod{q}} |L(1/2, \psi)|^4 \ll_\varepsilon q^{2+\varepsilon}$$

by a standard large sieve-type inequality. This suffices to finish the proof of Theorems 1.2 and 1.3 in the case that  $q$  is cube-free.

If  $q = p^3$  with  $p \equiv 1 \pmod{4}$  then (surprisingly!) there exist  $2(p-1)$  characters  $\psi$  modulo  $q$  such that  $|g(\chi, \psi)| = p^{1/2}q$ . These  $2(p-1)$  “singular” characters  $\psi$  form two cosets of the subgroup of characters modulo  $p$  sitting inside the group of all characters modulo  $p^3$ . So, we need to bound for two choices of  $\alpha$  primitive modulo  $p^3$  the sum

$$\sum_{\psi \pmod{p}} |L(1/2, \psi \cdot \alpha)|^4 g(\chi, \psi) \leq p^{\frac{1}{2}}q \sum_{\psi \pmod{p}} |L(1/2, \psi \cdot \alpha)|^4.$$

At this point, applying the Burgess bound individually to each  $L(1/2, \psi \cdot \alpha)$  gives  $\ll_{\varepsilon} q^{2+\varepsilon} p^{3/4}$ , while over-extending to all characters modulo  $p^3$  and using a large-sieve bound gives  $q^{2+\varepsilon} p^{1/2}$ . Neither of the bounds is sufficient. We would need a bound of the strength

$$(1.7) \quad \sum_{\psi \pmod{p}} |L(1/2, \psi \cdot \alpha)|^4 \ll_{\varepsilon} p^{2.5+\varepsilon}$$

for all  $\varepsilon > 0$ , which already gives a subconvex bound (though not even as strong as the Burgess bound), so one needs a treatment of moments of the rough shape (1.7) that goes beyond a large-sieve type inequality. We solve this problem of bounding fourth moments of Dirichlet  $L$ -functions along cosets by proving the following theorem.

**1.2. The fourth moment problem along subgroups.** Let  $T \geq 1$ , and  $q, d \geq 1$  be integers with  $d|q$ . Let  $q^* = \prod_{p \mid q} p^{\lceil \frac{2\beta}{3} \rceil}$ , so that  $q^*$  is the least positive integer so that  $q^2|(q^*)^3$ .

**Theorem 1.4.** *For all primitive  $\chi$  modulo  $q$  and  $\varepsilon > 0$  we have*

$$(1.8) \quad \int_{-T}^T \sum_{\psi \pmod{d}} |L(1/2 + it, \psi \cdot \chi)|^4 \ll_{\varepsilon} T \operatorname{lcm}(d, q^*) (qT)^{\varepsilon}.$$

Note that the set of characters  $\{\psi \cdot \chi : \psi \pmod{d}\}$  is a coset of the subgroup of characters modulo  $d$  inside the group of all characters modulo  $q$ . For example, if  $q = p^3$  and  $d = p^2$ , (1.8) is a Lindelöf-on-average upper bound, and more than suffices to establish the required estimate (1.7). This proves Theorems 1.2 and 1.3 in the case  $q = p^3$ .

In fact, Theorem 1.4 is strong enough to establish (1.6), and hence Theorems 1.2 and 1.3 in general. To see this, we perform an exhaustive calculation of the complete sums  $g(\chi, \psi)$  in Sections 2 and 3, culminating in Theorems 3.3 and 3.4. These two theorems form one of the main achievements of this paper, describing completely the structure of the cosets of singular characters  $\psi$  for which  $|g(\chi, \psi)|$  is exceptionally large.

Theorem 1.4 may be viewed as a  $q$ -aspect variant on Iwaniec’s [Iw1] short interval fourth moment bound

$$(1.9) \quad \int_T^{T+T^{2/3}} |\zeta(1/2 + it)|^4 dt \ll_{\varepsilon} T^{2/3+\varepsilon}.$$

See Section 1.5 below for more discussion on why these results are analogous. Iwaniec proves a number of other bounds on moments of zeta beyond (1.9), and it would be interesting to prove  $q$ -aspect variants of those bounds also. The second moment problem along cosets has been studied in some cases by Nunes [Nun] and recently by Milićević and White [MW].

There are many other works in the literature on different variants of the fourth moment problem for Dirichlet  $L$ -functions and the zeta function. To name just a few, we mention

[Iw1] [H-B2] [JM1] [Y1] [BM] [KMS] [BHKM]. Many of these papers focus on the problem of obtaining an *asymptotic* formula for the fourth moment, which leads to some difficult analytic problems that may be circumvented in the proof of Theorem 1.4, which is an upper bound. The asymptotic moment problem requires solving a shifted convolution sum in an unbalanced range, where the shift is very large compared to the length of summation. Since we are only interested in an upper bound, a simple Cauchy-Schwarz argument is able to completely sidestep this unbalanced problem (see Section 5).

It is also interesting to compare the subgroup structure of the family of Dirichlet characters appearing in (1.8) with the thin Galois orbits studied in [KMN].

**1.3. Shifted divisor sum with character.** The main problem faced in the proof of Theorem 1.4 is a strong bound on a shifted divisor sum with characters. We now discuss this problem. Suppose that  $w(x) = w_N(x)$  is a smooth weight function supported on  $x \asymp N$ . Let  $\chi$  be a primitive Dirichlet character modulo  $q$ . For  $h \geq 1$ , consider

$$(1.10) \quad \sum_n \chi(n+h)\tau(n+h)\bar{\chi}(n)\tau(n)w(n).$$

For analytic reasons, it is preferable to study a closely-related variant of the form

$$(1.11) \quad S(\chi, h) := \sum_{n \geq 1} \chi(n+h)\tau(n+h)\bar{\chi}(n) \sum_{n_1 n_2 = n} w(n_1, n_2),$$

where  $w(x, y)$  is smooth of compact support. One can always apply a partition of unity to write (1.10) as a short linear combination of sums of this form. We suppose  $w(x, y)$  is supported on  $x \asymp N_1$ ,  $y \asymp N_2$ , with  $N_1 N_2 = N$ . We also assume

$$h \ll N,$$

to avoid the more analytically difficult unbalanced shifted divisor sum.

We will also gain additional savings summing over  $h$ . Let

$$(1.12) \quad S(\chi) = \sum_{h \equiv 0 \pmod{d}} \sum_{n \geq 1} \chi(n+h)\tau(n+h)\bar{\chi}(n) \sum_{n_1 n_2 = n} w(n_1, n_2, h),$$

where  $w$  is part of a smooth family of functions of  $x, y, h$ , supported on  $x \asymp N_1$ ,  $y \asymp N_2$ , and  $h \ll H \ll N$ . The range relevant for proving Theorem 1.4 is  $N \ll q$ . We suppose that  $w$  satisfies

$$(1.13) \quad x^j y^k r^\ell w^{(j,k,\ell)}(x, y, r) \ll_{j,k,\ell} 1.$$

Note that we can view  $S(\chi)$  as a sum over  $h$  of  $S(\chi, h)$ , provided we allow  $w(x, y)$  appearing in (1.11) to also depend on  $h$ .

**Theorem 1.5.** *Suppose  $d|q$  and  $q^2|d^3$ . Then for all  $\varepsilon > 0$*

$$(1.14) \quad S(\chi) \ll_\varepsilon N \left(1 + \frac{H}{q}\right) (qN)^\varepsilon.$$

Applying an approximate functional equation and orthogonality of characters, Theorem 1.4 follows quickly from Theorem 1.5. The reduction step is detailed in Section 5.

A pleasant technical feature of Theorem 1.5 is that the bound in (1.14) does not include any factors that are sensitive to the current progress towards the Ramanujan conjecture (typically, the spectral analysis of shifted convolution sums with an individual shift will give rise to such dependence). The work of Blomer and Milićević [BM] also has this feature,

which in the proof arose from a clever arrangement of Hölder's inequality after the spectral decomposition of the shifted convolution sum, and we were able to adapt their idea to our present setting.

**1.4. A sketch.** Recall that  $q$  is our main parameter, and  $d$  is an integer with  $d|q$  and  $q^2|d^3$ . This hypothesis ensures that  $d$  and  $q$  share the same set of prime factors. In this sketch, we restrict ourselves to the special case  $q = p^3$  and  $d = p^2$ , which illustrates the nature of the argument in a relatively simple situation.

A main idea of the proof of Theorem 1.5 is that the sum  $S(\chi)$  exhibits a conductor dropping phenomenon: writing  $h = h'p^2$ , we have

$$(1.15) \quad \overline{\chi}(n)\chi(n+h) = \chi(1+h'\overline{n}p^2) = e_p(\ell_\chi h'\overline{n})$$

for some integer  $\ell_\chi$  with  $(\ell_\chi, p) = 1$ , since  $\chi$  has conductor  $p^3$ . Thus

$$(1.16) \quad S(\chi) \approx \sum_{h' \ll \frac{H}{p^2}} \sum_{n \asymp N} \tau(n+h'p^2)\tau(n)e_p(\ell_\chi h'\overline{n}).$$

(In this sketch, we use the symbol  $\approx$  merely to mean that the left hand side may be transformed into an expression resembling the right hand side, plus an acceptable error term.) Observe that in (1.16), there is no possible cancellation in the interior sum when  $p \mid h'$ . However, these terms make a contribution of at most  $\ll_\varepsilon N \frac{H}{q} p^\varepsilon$ , which is acceptable, so we may assume that  $(h', p) = 1$  from here on. This step corresponds to the factorization  $q = q_1 q_2$  in Corollary 8.6, i.e. in the present sketch we may assume that  $q = q_2 = p^3$  and  $h_{q_2} = p^2$ .

Next we solve the shifted convolution problem in (1.16). There are many ways to do this, and we opt to use an approximate functional equation-type formula for the divisor function of the rough form

$$(1.17) \quad \tau(n+h) \approx \sum_{c \ll \sqrt{N}} \frac{S(n+h, 0; c)}{c},$$

a method which is similar to using either the delta method or the circle method. After using (1.17) to separate  $n$  and  $h$ , we apply a double Poisson summation (i.e. Voronoi summation) to the sum over  $n$ .

It is technically more convenient not to use the formula (1.15) at the outset, and instead to first apply an approximate functional equation-type formula similar to (1.17) for  $\tau(n)\chi(n)$  (see Lemma 7.3). We then use the conductor dropping formula (1.15) in the course of computing the complete character sums that result from Poisson summation (see also the remarks following Corollary 8.7).

Either way, the result of these steps is a formula of the shape

$$(1.18) \quad S(\chi) \approx \sum_{\substack{c \ll \sqrt{N} \\ (c,p)=1}} \frac{N}{c^2 p^2} \sum_{\substack{h' \ll \frac{H}{p^2} \\ (h',p)=1}} \sum_{n_1 n_2 \ll \frac{(cp)^2}{N}} S(p^2 h', -\overline{p}^2 n_1 n_2; c) \text{Kl}_3(\ell_\chi h', \overline{c} n_1, \overline{c} n_2; p).$$

The formula (1.18) is a simplified form of (8.14). Note that the dual sum after Poisson summation is of length  $\frac{c^2 p^2}{N} \ll p^2$ , while the original length was of size  $N \ll q = p^3$ , so this represents a significant savings. At this point, if one uses the Weil bound for Kloosterman

sums and Deligne's bound for hyper-Kloosterman sums, we obtain only  $S(\chi) \ll pN^{3/4} \frac{H}{p^2} = H \frac{N^{3/4}}{p}$ , which is far from what is needed for Theorem 1.5 or even the weaker goal of (1.7).

To go further, we apply spectral methods from the theory of automorphic forms to the sum over  $c$  in the guise of the Bruggeman-Kuznetsov formula (see section 6). We first must resolve the  $\bar{c}$  inside the argument of the  $\text{Kl}_3$ , and do so by expanding into multiplicative characters, i.e. using the formula

$$\text{Kl}_3(x, y, z; p) = \frac{1}{\varphi(p)} \sum_{\eta \pmod{p}} \tau(\bar{\eta})^3 \eta(xyz),$$

when  $(xyz, p) = 1$  (see Lemma 8.8 for the general version). This leads to

$$(1.19) \quad S(\chi) \approx \frac{N}{p^3} \sum_{\eta \pmod{p}} \eta(\ell_\chi) \tau(\bar{\eta})^3 \sum_{h' \ll \frac{H}{p^2}} \sum_{n_1 n_2 \ll \frac{(cp)^2}{N}} \eta(h' n_1 n_2) \sum_{\substack{c \ll \sqrt{N} \\ (c,p)=1}} \frac{\bar{\eta}^2(c)}{c^2} S(ph', -\bar{p} n_1 n_2; c),$$

where we also used  $S(p^2 h', -\bar{p}^2 n_1 n_2; c) = S(ph', -\bar{p} n_1 n_2; c)$ . Now we may apply the Bruggeman-Kuznetsov formula for  $\Gamma_0(p)$  with central character  $\eta^2$  at the cusps  $\infty, 0$  to the sum over  $c$  in (1.19). After some careful analysis of test functions, we obtain a spectral reciprocity formula for  $S(\chi)$  of the rough shape

$$(1.20) \quad S(\chi) \approx \frac{N}{p^2} \sum_{\eta \pmod{p}} \eta(\ell_\chi) \frac{\tau(\bar{\eta})^3}{p^{3/2}} \sum_{t_j \ll 1} \sum_{\pi \in \mathcal{H}_{it_j}(p, \eta^2)} \epsilon(\pi)_{\text{fin}} \bar{\lambda}_\pi(p) L(1/2, \bar{\pi} \otimes \eta)^3 + (\text{Hol.}) + (\text{Eis.}),$$

where (Hol.) and (Eis.) represent similar contributions from holomorphic cusp forms and Eisenstein series, respectively, and  $\epsilon(\pi)_{\text{fin}}$  is the finite part of the root number of  $\pi$ . See (9.39) for the closest cousin to (1.20).

Applying Hölder's inequality, we are reduced to the problem of bounding

$$(1.21) \quad \sum_{\eta \pmod{p}} \sum_{t_j \ll 1} \sum_{\pi \in \mathcal{H}_{it_j}(p, \eta^2)} |L(1/2, \bar{\pi} \otimes \eta)|^4.$$

Using that  $\bar{\pi} \otimes \eta \in \mathcal{H}_{it_j}(p^2, 1)$ , we can bound this with a standard spectral large sieve inequality for level  $p^2$ . The restriction to  $q = p^3$  and  $d = p^2$  in this sketch has led to (1.21) being an overly-simplistic fourth moment problem. In Theorem 7.6 below, we bound the more general and difficult moment that arises. See the remarks following Theorem 7.6 for further discussion on this independently-interesting problem.

It is also instructive to compare the above sketch with Motohashi's spectral decomposition of the (smoothed) fourth moment of the zeta function over a short interval. Theorem 5.1 of [Mo] gives, roughly,

$$(1.22) \quad \int_T^{T+H} |\zeta(1/2 + it)|^4 dt \leftrightarrow (\text{main term}) + \frac{H}{\sqrt{T}} \sum_{t_j \ll \frac{T}{H}} L(1/2, u_j)^3 t_j^{-1/2} \sin(\theta_j),$$

where  $\theta_j \sim t_j \log t_j$ . Motohashi derives (1.22) from an exact formula for the weighted fourth moment of zeta, and the sequence of steps used in the proof is similar to that presented in the above sketch. In particular, the dual family of Maass cusp forms arises from a spectral decomposition of the shifted divisor problem. In Motohashi's case, the shifted divisor problem includes a  $t$ -aspect oscillatory factor, as in  $\sum_n \tau(n) \tau(n+h) \left(\frac{n+h}{n}\right)^{iT}$ ; this should be

compared with (1.10). The fact that  $h$  is small means that  $(\frac{n+h}{n})^{iT} \approx \exp(iT\frac{h}{n})$ , which is an archimedean analog of the conductor-dropping phenomenon of (1.15). It is pleasant to compare (1.22) with (1.20); taking  $H = T^{2/3}$  gives the closest comparison. The archimedean oscillatory factor  $\sin(\theta_j)$  is analogous to the argument of  $\tau(\bar{\eta})^3$ , which is in line with Stirling's approximation, and the analogy between Gauss sums and the gamma function.

**1.5. Remarks on close-knit families.** A key idea going into the proof of Theorem 1.1 is the shape of the family of automorphic forms in Theorems 1.2 and 1.3. This is yet another example of the by now well-known and powerful technique of deforming in a family of automorphic forms or  $L$ -functions (see [SST] for more discussion). To this end, we now offer some brief remarks on families in an ad-hoc context, which may be useful for interpreting the moment problems considered in this article and our previous work [PY2].

To fix ideas, let us work in the context of some ambient family of automorphic forms  $\mathcal{F}$ . Let  $\pi_0 \in \mathcal{F}$  and suppose that one wishes to prove subconvexity for  $L(\pi_0, 1/2)$ . A typical strategy is to choose a sub-family  $\mathcal{F}_0 \subseteq \mathcal{F}$  containing  $\pi_0$ , and consider, for example, a second moment of  $L$ -functions of the form  $\sum_{\pi \in \mathcal{F}_0} |L(\pi, 1/2)|^2$ . It is then advantageous to choose the family  $\mathcal{F}_0$  to have high spectral completeness, while at the same time to be as small as possible.

A natural way to quantify the closeness of two automorphic forms or representations is through the quantity

$$(1.23) \quad D(\pi_1, \pi_2) := \frac{C(\pi_1 \otimes \bar{\pi}_2)}{C(\pi_1 \otimes \bar{\pi}_1)^{1/2} C(\pi_2 \otimes \bar{\pi}_2)^{1/2}},$$

where  $C$  is the analytic conductor. Given a family  $\mathcal{F}_0$ , one can reasonably speak of the diameter of  $\mathcal{F}_0$  with respect to  $D(\pi_1, \pi_2)$ . Alternatively, one can define a sub-family  $\mathcal{F}_0 \subseteq \mathcal{F}$  by  $\mathcal{F}_0 = \mathcal{F}_0(r) = \{\pi \in \mathcal{F} : D(\pi, \pi_0) \leq r\}$ . Such small families fit into the framework of harmonic families of [SST], since the analytic conductor is a local invariant and Rankin-Selberg convolutions may be computed locally. Informally, we call families with small  $D(\pi_1, \pi_2)$  'close-knit'.

Working locally we can be a bit more precise. Let  $k$  be a non-archimedean local field with finite residue field. It follows easily from much more general work of Bushnell-Henniart [BH] (see [La, Thm. 1]) that the function

$$(1.24) \quad d(\pi_1, \pi_2) := c(\pi_1 \otimes \bar{\pi}_2) - \frac{1}{2}c(\pi_1 \otimes \bar{\pi}_1) - \frac{1}{2}c(\pi_2 \otimes \bar{\pi}_2),$$

where  $c$  is the conductor exponent, defines a pseudometric on the space of irreducible supercuspidal representations of  $\mathrm{GL}_n(k)$ .

We now consider some simple examples. Let  $\mathcal{F}$  be the set of Dirichlet characters and let  $\pi_0 = \chi \in \mathcal{F}$  be of conductor  $q$ . For  $d \mid q$ , the set  $\mathcal{F}_\chi(d) = \{\chi \cdot \psi : \psi \pmod{d}\}$  is an example of a close-knit family of diameter  $d$  around  $\chi$ . The family  $\mathcal{F}_\chi(d)$  is precisely the family considered in Theorem 1.4 (see also [Nun] and [MW]).

Considering the archimedean aspect, one finds many examples of families grouped according to  $D(\pi_1, \pi_2)$  in the literature. The short-interval  $t$ -aspect integral found in (1.9) is such an instance. To describe a slightly more advanced example, let  $\mathcal{F}$  be the set of Hecke-Maass eigenforms for  $\mathrm{SL}_2(\mathbb{Z})$ . Given  $u \in \mathcal{F}$ , write  $t_u$  for its spectral parameter. For parameters  $1 \ll \Delta \ll T$ , let  $\mathcal{F}_T(\Delta) = \{u \in \mathcal{F} : T < t_u \leq T + \Delta\}$ . The conductor of  $u \otimes u'$  is  $\asymp (1 + |t_u - t_{u'}|)^2 (1 + |t_u + t_{u'}|)^2 \ll \Delta^2 T^2$ , so that the close-knit family  $\mathcal{F}_T(\Delta)$  has diameter

$\ll \Delta^2$ . The cubic moment of  $L$ -functions over this family  $\sum_{u \in \mathcal{F}_T(\Delta)} L(1/2, u)^3$  was studied by Ivić [Iv], from which he derived Weyl-strength subconvexity in the spectral aspect.

The family of automorphic forms appearing in Theorem 1.2 provides another example. Let  $\mathcal{F} = \mathcal{H}_{it}(q^2, 1)$ . For  $\chi$  a primitive character modulo  $q$  that is *not* quadratic, consider the family of twists

$$(1.25) \quad \mathcal{F}_\chi := \{\pi \otimes \chi : \pi \in \mathcal{H}_{it}(m, \overline{\chi}^2), m \mid q\} \subseteq \mathcal{F}.$$

The family  $\mathcal{F}_\chi$  admits a simple interpretation in terms of local representation theory. The local components of  $\pi \otimes \chi \in \mathcal{F}_\chi$  are principal series at all finite places. Precisely,  $(\pi \otimes \chi)_p \simeq \pi(\chi_p, \overline{\chi}_p)$  for all  $p < \infty$ , where  $\chi_p$  is a quasi-character of  $\mathbb{Q}_p^\times$  whose restriction to  $\mathbb{Z}_p^\times$  matches the restriction of  $\chi$  to  $\mathbb{Z}_p^\times$ . Thus, the family  $\mathcal{F}_\chi$  could also have been described by specifying the local component at finitely many places of ramification to be a single principal series representation (up to unramified twists). Locally at  $p$ , we have  $d(\pi_{1,p}, \pi_{2,p}) = 0$  for any  $\pi_1, \pi_2 \in \mathcal{F}_\chi$ , so the family  $\mathcal{F}_\chi$  is as close-knit as possible at finite places.

Another interesting example occurs for thin Galois orbits of Dirichlet  $L$ -functions; see [KMN, pp. 6961-6963] for more details.

It is illuminating to view many families of  $L$ -functions under this lens, and the authors hope that this way of thinking may lead to beneficial choices of families of  $L$ -functions for problems in analytic number theory.

**1.6. Bounds on character sums.** Theorem 1.1 leads to an improvement on the Burgess bounds for character sums in some ranges.

**Theorem 1.6.** *For all primitive Dirichlet characters  $\chi$  modulo  $q$ ,  $x \geq 1$ , and  $\varepsilon > 0$  we have*

$$(1.26) \quad \sum_{n \leq x} \chi(n) \ll \begin{cases} x^{1/2} q^{11/64 + \varepsilon} \\ x^{8/15} q^{7/45 + \varepsilon}. \end{cases}$$

*Remarks.* The former bound is better than the latter for  $x \gg q^{47/96}$ . Recall the Burgess bound states  $\sum_{y < n \leq y+x} \chi(n) \ll x^{1-\frac{1}{r}} q^{\frac{r+1}{4r^2} + \varepsilon}$ , for  $r = 2, 3$ , and for any  $r \geq 1$  if  $q$  is cube-free (see [IK, Thm. 12.6]). Theorem 1.6 improves on the Burgess bounds with  $y = 0$  and  $r = 2$  or 3 in all non-trivial ranges.

*Sketch of proof.* Let  $0 < h < x$  be a parameter to be chosen later. Let  $w$  be a smooth weight function so that  $w(t) = 1$  for  $0 \leq t \leq x$ ,  $w(t) = 0$  for  $t \geq x+h$ , and satisfying  $w^{(j)}(t) \ll_j h^{-j}$ , for all  $t > 0$ . Then

$$S(\chi, w) := \sum_{n=1}^{\infty} \chi(n) w(n) = \frac{1}{2\pi i} \int_{(\sigma)} \tilde{w}(s) L(s, \chi) ds.$$

Integration by parts shows that the integral may be essentially truncated at  $\text{Im}(s) \ll x/h$ . Taking  $\sigma = 1/2$  and using Theorem 1.1 gives a bound on the smoothed sum, showing  $S(\chi, w) \ll x^{1/2} q^{1/6 + \varepsilon} (x/h)^{1/6}$ . Next, we have  $\sum_{n \leq x} \chi(n) = S(\chi, w) - \sum_{x < n \leq x+h} \chi(n) w(n)$ . For the latter sum, we may use summation by parts and the Burgess bound with  $r = 2$  or  $r = 3$ . Choosing  $h$  optimally then gives the two bounds.  $\square$

The interested reader may derive additional bounds for cube-free conductors using the Burgess bound for larger values of  $r$  in the final step of the above proof. The authors thank Roger Heath-Brown for suggesting the use of the Burgess bound on the short interval.



**1.7. Organization of the paper.** This paper is divided into two parts that are almost entirely independent of each other, and the notation is not necessarily consistent between the two parts. The authors believe this is a feature and not a bug.

The first part of this paper is devoted to the cubic moment problem and its reduction to the bound on the fourth moment along subgroups (i.e., Theorem 1.4), and is contained in Sections 2–4. Specifically, Section 2 contains a variety of character sum lemmas, Section 3 has a full analysis of  $g(\chi, \psi)$ , and Section 4 finishes the proof of Theorem 1.1 given the veracity of Theorem 1.4.

The second part gives the proof of Theorem 1.4, and is contained in Sections 5–9. Section 5 briefly deduces the proof from the shifted sum bound (Theorem 1.5). In Section 6, we present the background from the theory of automorphic forms, with an emphasis on the use of canonically-normalized Fourier expansions in the style of [MV]. Section 7 contains some tools from analytic number theory. The proof of the shifted sum bound begins in earnest in Section 8 and is completed in Section 9.

### 1.8. Notation.

- We denote by  $\mathbb{N}$  the set of natural numbers  $\{1, 2, \dots\}$  without zero.
- For a finite abelian group  $G$ , we denote by  $\widehat{G}$  its unitary dual. Exception: in Sections 6 and 9 we write  $\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p \simeq \varprojlim \mathbb{Z}/n\mathbb{Z}$ .
- For  $\pi$  a newform/automorphic representation on  $\mathrm{GL}_2$  and  $\chi$  a Dirichlet character, there are (at least) two standard conventions for the meaning of  $L(s, \pi \otimes \chi)$ . One convention is that it equals the straightforward Dirichlet series  $\sum_n \lambda_\pi(n) \chi(n) n^{-s}$ , and the other is that it equals the automorphic  $L$ -function associated to the twist of  $\pi$  by the Hecke character corresponding to  $\chi$ . In this paper, all  $L$ -functions of the form  $L(s, \pi \otimes \chi)$  use the automorphic definition. However, since the two conventions can only differ at Euler factors corresponding to primes dividing the conductor of  $\pi$  or the conductor of  $\chi$ , all statements of theorems or lemmas involving  $L(s, \pi \otimes \chi)$  remain equally valid using either convention.
- For  $\chi$  a Dirichlet character, we use  $L(s, \chi)$  to denote the classically-defined Dirichlet series  $\sum_n \chi(n) n^{-s}$ . If  $\chi$  is primitive, this agrees with the automorphic convention. If  $\chi$  is not primitive, but is induced by  $\chi^*$ , then it is easy to convert between  $L(s, \chi)$ , and  $L(s, \chi^*)$ .

**1.9. Acknowledgements.** The authors thank Roger Heath-Brown, Rizwan Khan, Emmanuel Kowalski, Djordje Milićević, and Lillian Pierce for comments and encouragement. We also thank the referees for many corrections and helpful suggestions.

## 2. CHARACTER SUMS TO PRIME POWER MODULUS

In this section we collect some lemmas that are useful for evaluating the character sums to prime-power modulus that arise in our work.

### 2.1. The Postnikov formula.

**Lemma 2.1.** *Let  $p$  be an odd prime, and  $\beta \geq 2$ . There exists a unique group homomorphism  $\ell : (\mathbb{Z}/p^\beta\mathbb{Z})^\times \rightarrow \mathbb{Z}/p^{\beta-1}\mathbb{Z}$ ,  $\chi \mapsto \ell_\chi$ , such that the Postnikov formula holds: for each Dirichlet character  $\chi$  modulo  $p^\beta$  and  $t \in \mathbb{Z}$  we have*

$$(2.1) \quad \chi(1 + pt) = e_{p^\beta}(\ell_\chi \log_p(1 + pt)).$$

The map  $\ell$  is surjective, and for  $1 \leq \alpha \leq \beta$  we have that  $\ell_{\chi_1} \equiv \ell_{\chi_2} \pmod{p^{\beta-\alpha}}$  if and only if  $\chi_1 \overline{\chi_2}$  is a character modulo  $p^\alpha$ .

*Proof.* For  $1 \leq \alpha \leq \beta$ , consider the reduction modulo  $p^\alpha$  map

$$(\mathbb{Z}/p^\beta\mathbb{Z})^\times \rightarrow (\mathbb{Z}/p^\alpha\mathbb{Z})^\times,$$

and denote its kernel by  $U_\alpha$ . Let  $e(x)$  be the continuous character of  $\mathbb{Q}_p$  agreeing with  $e^{2\pi i x}$  for  $x \in \mathbb{Q}$ , and let  $e_{p^\beta}(x) = e(p^{-\beta}x)$ . Let  $\log_p : 1 + p\mathbb{Z}_p \rightarrow p\mathbb{Z}_p$  be the  $p$ -adic logarithm defined by the convergent power series expansion

$$\log_p(1+x) = x - x^2/2 + x^3/3 \mp \dots$$

It is easy to check that  $\log_p(1 + p^\beta\mathbb{Z}_p) \subseteq p^\beta\mathbb{Z}_p$ , and in fact

$$(2.2) \quad \log_p(1 + p^\beta x) \equiv p^\beta x \pmod{p^{2\beta}},$$

since  $p$  is odd.

Consider the map  $f : U_1 \rightarrow S^1$  defined by

$$f : t \mapsto e_{p^\beta}(\log_p(t)).$$

The function  $f$  is well-defined by (2.2), and is a group homomorphism since  $\log_p(xy) = \log_p(x) + \log_p(y)$  for  $x, y \in 1 + p\mathbb{Z}_p$  (see e.g. [Neu, Prop. 5.5]). We claim that  $f$  has order  $p^{\beta-1}$  in  $\widehat{U}_1$ . Indeed, if  $t = 1 + px \in U_1$ , then we have  $f(t)^{p^{\beta-2}} = e_{p^2}(\log_p(t)) = e_p(x)$ , so  $f^{p^{\beta-2}}$  is not trivial in  $\widehat{U}_1$ , yet  $U_1$  has order  $p^{\beta-1}$ . Therefore  $\widehat{U}_1$  is cyclic and  $f$  is a generator. Define  $\ell_\chi$  to be the unique integer modulo  $p^{\beta-1}$  such that  $\chi|_{U_1} = f^{\ell_\chi}$ , which is equivalent to the Postnikov formula (2.1). We easily see that  $\ell$  is a group homomorphism. Next we show this map is surjective. The kernel of  $\ell$  is the subgroup of characters trivial on  $U_1$ , which is isomorphic to  $(\widehat{\mathbb{Z}/p\mathbb{Z}})^\times$ . Hence by comparing cardinalities, we see  $\ell$  is surjective.

We claim that  $f|_{U_\alpha}$  has order  $p^{\beta-\alpha}$  in the group  $\widehat{U}_\alpha$ . Indeed, writing  $t = 1 + p^\alpha x$ , we have  $f(t)^{p^{\beta-\alpha-1}} = e_{p^\beta}(p^{\beta-\alpha-1} \log_p(t)) = e_{p^{\alpha+1}}(\log_p(t)) = e_p(x)$ , showing the claim. Then  $\chi|_{U_\alpha} = f|_{U_\alpha}^{\ell_\chi}$ , and we deduce that  $\ell_\chi \equiv 0 \pmod{p^{\beta-\alpha}}$  if and only if  $\chi|_{U_\alpha} = 1$ , which in turn is equivalent to the condition that  $\chi$  is a character modulo  $p^\alpha$ . The final statement of the lemma now follows, since  $\ell$  is a group homomorphism.  $\square$

**2.2. Character sums.** A rational function  $f \in \mathbb{Z}(t)$  is an equivalence class of pairs of polynomials  $f_1/f_2$  with integer coefficients and  $f_2$  not identically zero. An integer  $t_0$  is said to be in the domain of  $f$  if  $f_2(t_0) \neq 0$  with  $f = f_1/f_2$  written in lowest terms (i.e., with  $f_1$  and  $f_2$  coprime). Meanwhile, a rational function  $f \in (\mathbb{Z}/p^\beta\mathbb{Z})(t)$  is an equivalence class of pairs of polynomials  $f_1/f_2$  with coefficients in  $\mathbb{Z}/p^\beta\mathbb{Z}$  and with  $p$  not dividing all of the coefficients of  $f_2$ . Similarly,  $t_0 \in \mathbb{Z}/p^\beta\mathbb{Z}$  is said to be in the domain of  $f$  if  $p \nmid f_2(t_0)$  with  $f = f_1/f_2$  in lowest terms. (Recall that in a commutative ring  $A$ , two elements  $a, b \in A$  are called coprime if  $(a) + (b) = A$ .) If  $p$  does not divide  $x \in \mathbb{Z}/p^\beta\mathbb{Z}$  then we call  $x$  a “ $p$ -adic unit”. The above notions also extend naturally to several variables. Lastly, in the character sums of the form  $\sum_t^* \chi(f(t))\psi(g(t))$  that we study in Sections 2 and 3 of this paper, the  $*$  is always taken to mean that we sum over those  $t$  lying in the intersection of the domains of  $f$  and  $g$ .

Let  $g \in \mathbb{Z}(t)$  be a rational function whose reduction  $\bar{g}$  modulo  $p^\beta$  exists. Let  $t_0$  be an integer whose reduction modulo  $p^\beta$  lies in the domain of  $\bar{g}$ . Then, it is easy to see that

$\frac{g^{(n)}(t_0)}{n!} \in \mathbb{Z}_p$  for all  $n \geq 0$ . In particular, this shows that

$$(2.3) \quad g(x_0 + p^\beta x_1) \equiv g(x_0) + p^\beta g'(x_0)x_1 \pmod{p^{2\beta}}$$

for any integer  $x_0$  reducing to the domain of  $\bar{g}$ .

More generally, suppose that  $p$  does not divide the whole denominator of  $g \in \mathbb{Z}(t_1, \dots, t_n)$  and  $x_0 \in \mathbb{Z}^n$  reduces modulo  $p^\beta$  to lie in the domain of  $\bar{g}$ . Then, the Taylor expansion of  $g$  at  $x_0$  has coefficients in  $\mathbb{Z}_p$  and we have

$$(2.4) \quad g(x_0 + p^\beta x_1) \equiv g(x_0) + p^\beta g'(x_0)x_1 + p^{2\beta} \frac{1}{2} g''(x_0)[x_1] \pmod{p^{3\beta}},$$

where  $g'$  denotes the gradient of  $g$ ,  $g''$  is the Hessian matrix, and  $A[x] = x^T A x$  is the quadratic form associated to a square matrix  $A$  and evaluated at  $x$ .

For  $f_i \in \mathbb{Z}(t_1, \dots, t_n)$ ,  $i = 1, \dots, d$ , let  $f = (f_1, \dots, f_d) \in \mathbb{Z}(t_1, \dots, t_n)^d$  be the associated  $d$ -tuple of rational functions. For such an  $f$  we have the associated  $d \times n$  Jacobian matrix, which we denote by  $f' \in M_{d \times n}(\mathbb{Z}(t_1, \dots, t_n))$ . Similarly, we have the logarithmic Jacobian  $(\log f)'$ , where the  $ij$  entry is given by  $\partial_j f_i / f_i$ .

Define an additive character  $\theta$  modulo  $q = (q_1, \dots, q_d) \in \mathbb{N}^d$  as a group homomorphism  $\mathbb{Z}^d / q\mathbb{Z}^d \rightarrow \mathbb{C}^\times$ , lifted to  $\mathbb{Z}^d$  by periodicity. By the Chinese remainder theorem,  $\theta$  can be expressed uniquely as  $\theta(n) = \theta_1(n_1) \dots \theta_d(n_d)$ , with  $\theta_i(n) = e_{q_i}(a_{\theta_i} n)$  for some  $a_{\theta_i} \in \mathbb{Z}$ . If  $q$  is diagonal, we may abuse notation and write simply  $\theta(n) = e_q(a_\theta n)$  where  $a_\theta n$  is the standard scalar product.

Likewise, a Dirichlet character modulo  $q = (q_1, \dots, q_d)$  is a map  $(\mathbb{Z}^d / q\mathbb{Z}^d)^\times \rightarrow \mathbb{C}^\times$  extended to  $\mathbb{Z}^d$  in the natural way. Again,  $\chi$  may be expressed uniquely as  $\chi((n_1, \dots, n_d)) = \chi_1(n_1) \dots \chi_d(n_d)$ , where  $\chi_i$  is modulo  $q_i$ ,  $i = 1, \dots, d$ . If  $p$  is odd,  $q = (p^\beta, \dots, p^\beta)$  with  $\beta \geq 2$ , we define  $\ell_\chi = (\ell_{\chi_1}, \dots, \ell_{\chi_d})$  with  $\ell_{\chi_i}$  as in Lemma 2.1. Note that the Postnikov formula generalizes to give for  $n = (n_1, \dots, n_d)$  with each  $n_i \equiv 1 \pmod{p}$  the formula  $\chi(n) = e_{p^\beta}(\ell_\chi \log_p(n))$ , with the standard scalar product and where  $\log_p(n) = (\log_p(n_1), \dots, \log_p(n_d))$ .

**Lemma 2.2.** *Let  $p$  be an odd prime,  $\chi$  be a Dirichlet character modulo  $(p^{2\alpha}, \dots, p^{2\alpha})$ ,  $\theta$  be an additive character modulo  $(p^{2\alpha}, \dots, p^{2\alpha})$  and  $f, g \in \mathbb{Z}(t_1, \dots, t_n)^d$  as above. Consider the congruence*

$$(2.5) \quad \ell_\chi(\log f)'(t_0) + a_\theta g'(t_0) \equiv 0 \pmod{p^\alpha}.$$

We have

$$(2.6) \quad S := \sum_{t \in (\mathbb{Z}/p^{2\alpha}\mathbb{Z})^n}^* \chi(f(t))\theta(g(t)) = p^{n\alpha} \sum_{\substack{t_0 \in (\mathbb{Z}/p^\alpha\mathbb{Z})^n \\ (2.5) \text{ holds}}}^* \chi(f(t_0))\theta(g(t_0)).$$

The right hand side does not depend on the choice of lift of  $t_0$  to  $\mathbb{Z}_p^d$ .

Remark. This is a natural multi-variable generalization of [IK, Lem. 12.2].

*Proof.* Write  $t = t_0 + p^\alpha t_1$ , and  $\chi(f(t)) = \chi(f(t_0))\chi(f(t)/f(t_0))$ . Then, by the Postnikov formula (2.1), (2.2), and (2.4), we have

$$\chi(f(t)/f(t_0)) = e_{p^{2\alpha}}(\ell_\chi \log_p(f(t)/f(t_0))) = e_{p^\alpha}(\ell_\chi(\log_p f)'(t_0)t_1).$$

Similarly,  $\theta(g(t)) = \theta(g(t_0))\theta(g(t) - g(t_0))$ , and

$$\theta(g(t) - g(t_0)) = e_{p^\alpha}(a_\theta g'(t_0)t_1)$$

Then

$$S = \sum_{t_0 \pmod{p^\alpha}}^* \chi(f(t_0))\theta(g(t_0)) \sum_{t_1 \pmod{p^\alpha}} e_{p^\alpha}(\ell_\chi(\log f)'(t_0)t_1 + a_\theta g'(t_0)t_1).$$

The inner sum vanishes unless (2.5) holds, giving the formula stated in the lemma. The proof shows that the right hand side of (2.6) is independent of choice of lifts.  $\square$

Next we generalize the odd exponent case of [IK, Lem. 12.3]. To this end, we introduce multi-variable Gauss sums. Let  $L : \mathbb{Z}^n \rightarrow \mathbb{Z}$  be a linear form with integer coefficients, and  $Q : \mathbb{Z}^n \rightarrow \mathbb{Z}$  be a quadratic form (see e.g. [Se, Ch.IV Def. 1]). Define

$$(2.7) \quad G_p(Q, L) = \sum_{t \in \mathbb{F}_p^n} e_p(Q[t] + Lt).$$

**Lemma 2.3.** *Let  $p$  be an odd prime,  $\chi$  be a Dirichlet character modulo  $(p^{2\alpha+1}, \dots, p^{2\alpha+1})$ ,  $\theta$  be an additive character modulo  $(p^{2\alpha+1}, \dots, p^{2\alpha+1})$ , and  $f, g \in \mathbb{Z}(t_1, \dots, t_n)^d$  as above. Then*

$$S = \sum_{t \in (\mathbb{Z}/p^{2\alpha+1}\mathbb{Z})^n}^* \chi(f(t))\theta(g(t)) = p^{n\alpha} \sum_{\substack{t_0 \in (\mathbb{Z}/p^\alpha\mathbb{Z})^n \\ (2.5) \text{ holds}}}^* \chi(f(t_0))\theta(g(t_0))G_p(Q, L),$$

where

$$(2.8) \quad L = p^{-\alpha}(\ell_\chi(\log f)'(t_0) + a_\theta g'(t_0))$$

and  $Q$  is the quadratic form with associated matrix (in the standard basis for  $\mathbb{Z}^n$ ) given by

$$(2.9) \quad Q = \frac{1}{2}\ell_\chi(\log f)''(t_0) + \frac{1}{2}g''(t_0).$$

The right hand side does not depend on the choice of lift of  $t_0$  to  $\mathbb{Z}_p^d$ .

*Proof.* Write  $t = t_0 + p^\alpha t_1$ , and  $\chi(f(t)) = \chi(f(t_0))\chi(f(t)/f(t_0))$ . Then, by the Postnikov formula (2.1), (2.2), and (2.4), we have

$$\chi(f(t)/f(t_0)) = e_{p^{2\alpha+1}}(\ell_\chi \log_p(f(t)/f(t_0))) = e_{p^{2\alpha+1}}(\ell_\chi(\log f)'(t_0)t_1)e_p(\frac{1}{2}\ell_\chi(\log f)''(t_0)[t_1]).$$

Similarly,  $\theta(g(t)) = \theta(g(t_0))\theta(g(t) - g(t_0))$ , and

$$\theta(g(t) - g(t_0)) = e_{p^{2\alpha+1}}(a_\theta g'(t_0)t_1)e_p(\frac{1}{2}g''(t_0)[t_1]).$$

Changing variables  $t_1 \rightarrow t_1 + pe_j$ , where  $e_j$  is the  $j$ -th standard basis vector, leaves the quadratic terms unchanged. Hence the inner sum vanishes unless (2.5) holds, in which case we obtain the claimed result.  $\square$

In view of Lemma 2.3, it will be useful to estimate quadratic Gauss sums.

**Lemma 2.4.** *Let  $p$  be an odd prime, let  $Q$  be a quadratic form over  $\mathbb{F}_p$ , and  $L$  a linear form, as above. Let  $V$  be the isotropic subspace of  $Q$ . Let  $r_Q$  denote the rank of  $Q$ . Then  $G_p(Q, L)$  vanishes unless  $L|_V = 0$ , in which case*

$$|G_p(Q, L)| = p^{\frac{r_Q}{2}} p^{(n-r_Q)}.$$

*Proof.* It is well-known that one can change basis for  $\mathbb{F}_p^n$  so that the quadratic form  $Q$  is orthogonal with respect to this basis (e.g. see [Se, Ch.IV.1.4 Thm. 1]). In particular, we have  $\mathbb{F}_p^n = V \oplus U$  where  $V$  is the isotropic subspace of  $Q$ , and  $U$  is a complementary subspace.

Therefore, if  $v \in V$  and  $u \in U$ , then  $Q[v+u] = Q[u]$ . Using this basis to calculate the Gauss sum, we have

$$G_p(Q, L) = \left( \sum_{v \in V} e_p(Lv) \right) \left( \sum_{u \in U} e_p(Q[u] + Lu) \right).$$

Note that the sum over  $v$  vanishes unless  $L|_V = 0$ , while the sum over  $u$  has absolute value  $p^{r_Q/2}$ , where  $r_Q$  is the rank of the quadratic form, since  $U$  has a basis on which  $Q$  is diagonalized, and by the standard one-variable evaluation of quadratic Gauss sums. This completes the proof.  $\square$

Motivated by an application (namely, Lemma 2.8), we wish to mildly generalize Lemmas 2.2 and 2.3 as follows. Let  $p$  be an odd prime and suppose  $1 \leq \beta \leq \gamma$ . Let  $f, g \in \mathbb{Z}(t_1, \dots, t_n)$  with  $p$  not dividing every coefficient of the denominators of  $f, g$ . Let  $V$  be the subset of  $x \in (\mathbb{Z}/p^\beta\mathbb{Z})^n$  for which  $p^{\gamma-\beta}x$  modulo  $p^\gamma$  lies in the domain of  $f$  modulo  $p^\gamma$ . Then  $F(x) = f(p^{\gamma-\beta}x)$  defines a function  $F : V \rightarrow \mathbb{Z}/p^\gamma\mathbb{Z}$  and we call  $V$  the domain of  $F$ . Let us write  $G$  for the same construction applied to  $g$ .

These definitions extend component-wise, as follows. Let  $\gamma = (\gamma_1, \dots, \gamma_d)$  with each  $\gamma_i \geq \beta \geq 1$ , and  $p^\gamma = (p^{\gamma_1}, \dots, p^{\gamma_d})$ . Given  $f = (f_1, \dots, f_d), g = (g_1, \dots, g_d) \in \mathbb{Z}(x_1, \dots, x_n)^d$ , define  $F = (F_1, \dots, F_d)$  and  $G = (G_1, \dots, G_d)$  by  $F_i(x) = f_i(p^{\gamma_i-\beta}x)$  and  $G_i(x) = g_i(p^{\gamma_i-\beta}x)$  for all  $1 \leq i \leq d$ . Then  $F, G$  define functions with domains given by the intersection of the domains of the  $F_i, G_i$ , as above. Let  $\chi$  be a Dirichlet character modulo  $p^\gamma$  and  $\theta$  an additive character modulo  $p^\gamma$ . If  $F$  and  $G$  are two such  $d$ -tuples of rational functions, then the functions  $\chi(F(t))$  and  $\theta(G(t))$  are well-defined on the domains of  $F$  and  $G$ .

**Lemma 2.5.** *Let  $\gamma = (\gamma_1, \dots, \gamma_d)$  with each  $\gamma_i \geq 2\alpha \geq 2$ . Write  $\beta = 2\alpha$ . Let  $p$  be an odd prime,  $\chi$  a Dirichlet character modulo  $(p^{\gamma_1}, \dots, p^{\gamma_d})$ ,  $\theta$  an additive character modulo  $(p^{\gamma_1}, \dots, p^{\gamma_d})$ , and  $F, G$  as above. Define the congruence condition*

$$(2.10) \quad \ell_\chi(\log f)'(p^{\gamma-\beta}t_0) + a_\theta g'(p^{\gamma-\beta}t_0) \equiv 0 \pmod{p^\alpha}.$$

We have

$$(2.11) \quad S := \sum_{t \in (\mathbb{Z}/p^\beta\mathbb{Z})^n}^* \chi(F(t))\theta(G(t)) = p^{n\alpha} \sum_{\substack{t_0 \in (\mathbb{Z}/p^\alpha\mathbb{Z})^n \\ (2.10) \text{ holds}}}^* \chi(F(t_0))\theta(G(t_0)).$$

*Proof.* The proof is very similar to that of Lemma 2.2. We have

$$\chi(F(t_0 + p^\alpha t_1)) = \chi(F(t_0))\chi\left(\frac{F(t_0 + p^\alpha t_1)}{F(t_0)}\right) = \chi(F(t_0))e_{p^\gamma}\left(\ell_\chi \frac{F'}{F}(t_0)p^\alpha t_1\right),$$

and similarly

$$\theta(G(t_0 + p^\alpha t_1)) = \theta(G(t_0))\theta(a_\theta G'(t_0)p^\alpha t_1).$$

Therefore,

$$S = p^{n\alpha} \sum_{\substack{t_0 \pmod{p^\alpha} \\ (2.12) \text{ holds}}}^* \chi(F(t_0))\theta(G(t_0)),$$

where (2.12) is the congruence condition

$$(2.12) \quad \ell_\chi \frac{F'}{F}(t_0)p^\alpha t_1 + a_\theta G'(t_0)p^\alpha t_1 \equiv 0 \pmod{p^\gamma}.$$

Note that  $G'(t_0) = p^{\gamma-\beta} f'(p^{\gamma-\beta}t_0)$ , and likewise  $\frac{F'}{F}(t_0) = p^{\gamma-\beta}(\log f)'(p^{\gamma-\beta}t_0)$ . Hence the congruence condition (2.12) is seen to be the same as (2.10).  $\square$

Similarly, the generalization of Lemma 2.3 is given by:

**Lemma 2.6.** *Let  $\gamma = (\gamma_1, \dots, \gamma_d)$  with each  $\gamma_i \geq \alpha \geq 1$ . Write  $\beta = 2\alpha + 1$ . Let  $p$  be an odd prime,  $\chi$  a Dirichlet character modulo  $(p^{\gamma_1}, \dots, p^{\gamma_d})$ ,  $\theta$  an additive character modulo  $(p^{\gamma_1}, \dots, p^{\gamma_d})$ , and  $F, G$  as above. We have*

$$(2.13) \quad S := \sum_{t \in (\mathbb{Z}/p^\beta \mathbb{Z})^n}^* \chi(F(t))\theta(G(t)) = p^{n\alpha} \sum_{\substack{t_0 \in (\mathbb{Z}/p^\alpha \mathbb{Z})^n \\ (2.10) \text{ holds}}}^* \chi(F(t_0))\theta(G(t_0))G_p(Q, L),$$

where

$$L = p^{-\alpha}(\ell_\chi(\log f)'(p^{\gamma-\beta}t_0) + a_\theta g'(p^{\gamma-\beta}t_0)),$$

and  $Q$  is the quadratic form with associated matrix (in the standard basis for  $\mathbb{Z}^n$ ) given by

$$Q = \frac{1}{2}\ell_\chi p^{\gamma-\beta}(\log f)''(p^{\gamma-\beta}t_0) + \frac{1}{2}p^{\gamma-\beta}g''(p^{\gamma-\beta}t_0).$$

Since the proof is similar to those of Lemmas 2.3 and 2.5, we omit the details. For the sake of clarity, we remark that  $\ell_\chi p^{\gamma-\beta}(\log f)''(p^{\gamma-\beta}t_0)$  is shorthand for

$$\sum_{i=1}^d \ell_{\chi_i} p^{\gamma_i-\beta}(\log f_i)''(p^{\gamma_i-\beta}t_0),$$

and similarly for  $g''$ . It will be useful later, in the proof of Lemma 2.8, to observe that if  $\gamma_i > \beta$  then the  $i$ -th component makes no contribution to the quadratic form  $Q$ .

The following lemma, with its easy proof omitted, will be helpful for solving the linear congruence in (2.5) in future applications.

**Lemma 2.7.** *Let  $R$  be a commutative ring, with group of units  $R^\times$ . Let  $M = (a_{ij}) \in M_{2 \times 2}(R)$  with  $a_{ij} \in R^\times$  for all  $i, j$ . Then there is a solution to  $(x_1, x_2).M = (0, 0)$  with  $x_1, x_2 \in R^\times$  if and only if  $\det(M) = 0$ , in which case the solutions are given by  $x_1 a_{11} + x_2 a_{21} = 0$  (whence  $x_1 = a_{21}r$ ,  $x_2 = -a_{11}r$ , for some  $r \in R^\times$ ).*

**2.3. Application.** In [PY2, Conj. 6.6], we left the estimation of a certain character sum as a conjecture, which we prove here aided by Lemmas 2.5 and 2.6.

**Lemma 2.8.** *Let  $p$  be an odd prime, let  $\chi$  be a multiplicative character of conductor  $p^\gamma$ ,  $\gamma \geq 2$ , and suppose  $\psi$  is a multiplicative character with conductor  $p^\beta$ ,  $1 \leq \beta < \gamma$ . Then*

$$(2.14) \quad \sum_{y \pmod{p^\beta}} \sum_{u \pmod{p^\beta}} \bar{\chi}(up^{2(\gamma-\beta)}y + 1)\chi(1 + p^{\gamma-\beta}y)\chi(1 - p^{\gamma-\beta}u)\psi(u)\psi(y) \ll p^\beta.$$

*Proof.* This is an instance of  $S$  defined by (2.11), (2.13), where  $d = 2$ ,  $\gamma_1 = \gamma$ ,  $\gamma_2 = \beta$ ,

$$(f_1(y, u), f_2(y, u)) = \left( \frac{(1+y)(1-u)}{1+yu}, yu \right), \quad (\chi_1, \chi_2) = (\chi, \psi),$$

$F_1(y, u) = f_1(p^{\gamma-\beta}y, p^{\gamma-\beta}u)$ ,  $F_2(y, u) = f_2(y, u)$ , and of course the additive character is not present. A short calculation shows

$$(\log f)'(p^{\gamma-\beta}t) = \left( \frac{1-p^{\gamma-\beta}u}{(1+p^{\gamma-\beta}y)(1+p^{2(\gamma-\beta)}yu)} \frac{-1-p^{\gamma-\beta}y}{u^{-1}}, \frac{-1-p^{\gamma-\beta}y}{u^{-1}} \right).$$

The summations in Lemmas 2.5 and 2.6 run over  $t_0$  such that

$$\ell_\chi(\log f)'(p^{\gamma-\beta}t_0) \equiv 0 \pmod{p^\alpha},$$

where  $\alpha = \beta/2$  for  $\beta$  even and  $\alpha = (\beta - 1)/2$  for  $\beta \geq 3$  odd, so we write  $t_0 = (u_0, y_0)$  and work out what this means in terms of conditions on  $u_0$  and  $y_0$ . Some simple algebra (cf. Lemma 2.7) shows that this reduces to  $u_0 \equiv -y_0 \pmod{p^\alpha}$ , which uniquely determines  $u_0$  in terms of  $y_0$ , and then

$$(2.15) \quad \frac{\ell_\psi}{y_0} \equiv \frac{-\ell_\chi}{1 - p^{2(\gamma-\beta)}y_0^2} \pmod{p^\alpha},$$

which uniquely determines  $y_0 \pmod{p^\alpha}$ , by Hensel's lemma. Hence, when  $\beta$  is even,  $|S| \leq p^\beta$ , by Lemma 2.5, giving the bound (2.14).

Now consider the case that  $\beta = 2\alpha + 1 \geq 3$  is odd; it was already shown above that  $u_0$  and  $y_0$  are uniquely determined modulo  $p^\alpha$ , so the only remaining question is the size of the Gauss sum  $G_p(Q, L)$ . It is easy to see that  $Q$  is non-singular, since only the  $f_2$ -aspect enters into the calculation, and the Hessian of  $\log f_2$  is diagonal with entries  $-y_0^{-2}, -u_0^{-2}$ . Therefore,  $|G_p(Q, L)| = p$ , and (2.14) follows immediately.

Finally, we consider the case  $\beta = 1$ . In this case, we have  $\chi(1 + p^{\gamma-1}x) = e_p(\ell_\chi x)$ , for any  $x \in \mathbb{Z}$ , so it is easy to directly evaluate (2.14) as a product of two Gauss sums, giving the desired bound.  $\square$

**2.4. The case  $p = 2$ .** The previous work in this section largely assumed  $p \neq 2$ . The case  $p = 2$  has some minor differences, and for clarity we treat this case separately.

**Lemma 2.9.** *Let  $p = 2$ , and  $\beta \geq 3$ . There exists a unique group homomorphism  $\ell : (\widehat{\mathbb{Z}/p^\beta\mathbb{Z}})^\times \rightarrow \mathbb{Z}/p^{\beta-2}\mathbb{Z}$ ,  $\chi \mapsto \ell_\chi$ , such that the Postnikov formula holds: for each Dirichlet character  $\chi$  modulo  $p^\beta$  and  $t \in \mathbb{Z}$  we have*

$$(2.16) \quad \chi(1 + 4t) = e_{p^\beta}(\ell_\chi \log_2(1 + 4t)).$$

*The map  $\ell$  is surjective, and for  $2 \leq \alpha \leq \beta$  we have that  $\ell_{\chi_1} \equiv \ell_{\chi_2} \pmod{2^{\beta-\alpha}}$  if and only if  $\chi_1 \overline{\chi_2}$  is a character modulo  $2^\alpha$ .*

The proof is very similar to the case  $p > 2$ , so we give only a brief outline of the proof. Using the notation  $U_\alpha$  from the proof of Lemma 2.1, define  $f : U_2 \rightarrow S^1$  by  $f(t) = e_{2^\beta}(\log_p(t))$ . One easily checks that  $f$  is well-defined and has order  $2^{\beta-2}$ , so  $\widehat{U}_2$  is cyclic generated by  $f$ . Therefore, (2.16) holds for some  $\ell_\chi$ . The final statement of the lemma is easy to check.

**Lemma 2.10.** *Let  $p = 2$ . Let  $\chi$  be a Dirichlet character modulo  $p^\beta$ , and  $\psi$  be an additive character modulo  $p^\beta$ , where  $\beta \geq 3$ . Let  $f, g \in \mathbb{Z}(t_1, \dots, t_n)^d$  as in Lemma 2.2. Let  $\alpha = \lfloor \frac{\beta-1}{2} \rfloor$ . Then*

$$(2.17) \quad S := \sum_{t \in (\mathbb{Z}/p^\beta\mathbb{Z})^n}^* \chi(f(t))\psi(g(t)) = p^{n\alpha} \sum_{\substack{t_0 \in (\mathbb{Z}/p^{\beta-\alpha}\mathbb{Z})^n \\ \ell_\chi(\log f)'(t_0) + a_\psi g'(t_0) \equiv 0 \pmod{p^\alpha}}}^* \chi(f(t_0))\psi(g(t_0)),$$

*where the star indicates that the sum runs over numbers for which  $f_i(t) \in \mathbb{Z}_p^\times, g_i(t) \in \mathbb{Z}_p$ . The right hand side does not depend on the choice of lifts of  $f(t_0)$  and  $g(t_0)$  to  $\mathbb{Z}_p^d$ .*

**Remark.** For our later purposes, this result is a suitable replacement for Lemmas 2.2 and 2.3. In practice, the linear congruence almost entirely determines  $t_0 \pmod{p^\alpha}$  (which then almost entirely determines  $t_0 \pmod{p^{\beta-\alpha}}$ , since  $\beta - \alpha = \alpha + O(1)$ , and  $p^{O(1)} = O(1)$  for  $p = 2$ ).

*Proof.* Let  $t = t_0 + 2^{\beta-\alpha}t_1$ . Since  $\beta \geq 3$ , we have  $\alpha \leq \beta - 2$ , so  $\frac{f(t)}{f(t_0)} \equiv 1 \pmod{4}$ . Then

$$\chi(f(t)) = \chi(f(t_0))\chi(f(t)/f(t_0)) = \chi(f(t_0))e_{2^\beta}(\ell_\chi \log_2(f(t)/f(t_0))).$$

Next we note

$$\log_2(f(t)/f(t_0)) \equiv (\log f)'(t_0)2^{\beta-\alpha}t_1 \pmod{2^\beta},$$

under the assumption  $2(\beta - \alpha) - 1 \geq \beta$ , equivalently,  $\alpha \leq \frac{\beta-1}{2}$ . Note  $\frac{\beta-1}{2} \leq \beta - 2$  since  $\beta \geq 3$ . The rest of the proof then proceeds exactly as in Lemma 2.2.  $\square$

**Lemma 2.11.** *The bound in Lemma 2.8 holds for  $p = 2$ .*

The proof is similar to the odd  $p$  case, so we omit the details.

### 3. THE BEHAVIOR OF $g(\chi, \psi)$

**3.1. Introductory lemmas.** Let  $A \in \mathbb{Z}$ , and let

$$Q(x) = Q_A(x) = x^2 + Ax - 1 \in \mathbb{Z}[x].$$

For an odd prime  $p$  and integer  $\beta \geq 1$ , define

$$r(A, p^\beta) = \#\{x \pmod{p^\beta} : Q_A(x) \equiv 0 \pmod{p^\beta}\}.$$

Let  $\Delta = A^2 + 4$  be the discriminant of  $Q_A$ . By completing the square, note

$$(3.1) \quad Q(x) = \left(x + \frac{A}{2}\right)^2 - \frac{\Delta}{4}.$$

We then have  $r(A, p^\beta) = \rho(\Delta, p^\beta)$ , where

$$\rho(\Delta, p^\beta) := \#\{x \pmod{p^\beta} : x^2 \equiv \Delta \pmod{p^\beta}\}.$$

**Lemma 3.1.** *Let  $p$  be an odd prime, and  $\beta \geq 1$ . If  $p \nmid \Delta$ , then*

$$\rho(\Delta, p^\beta) = \rho(\Delta, p) = 1 + \left(\frac{\Delta}{p}\right).$$

*If  $p^\beta | \Delta$ , then*

$$\rho(\Delta, p^\beta) = p^{\lfloor \beta/2 \rfloor} = \begin{cases} p^{\frac{\beta}{2}}, & \beta \text{ even,} \\ p^{\frac{\beta-1}{2}}, & \beta \text{ odd.} \end{cases}$$

*If  $p | \Delta$ , but  $p^\beta \nmid \Delta$ , then*

$$(3.2) \quad \rho(\Delta, p^\beta) \leq 2(\Delta, p^\beta)^{1/2} \delta((p^\beta, \Delta) = \square).$$

*Proof.* The case  $p \nmid \Delta$  follows from Hensel's lemma. The conclusion when  $p^\beta | \Delta$  is easy to verify directly.

Now suppose  $(\Delta, p^\beta) = p^\alpha$ , with  $1 \leq \alpha < \beta$ . Write  $\Delta = p^\alpha \Delta'$  with  $(\Delta', p) = 1$ . It is easy to see that if  $\alpha$  is odd then  $\rho(\Delta, p^\beta) = 0$ . If  $\alpha$  is even (which means  $(p^\beta, \Delta) = \square$ ) then we write  $x = p^{\alpha/2}x_1$ , say, where  $x_1$  runs modulo  $p^{\beta-\frac{\alpha}{2}}$ . Then  $x_1$  solves the congruence

$$x_1^2 \equiv \Delta' \pmod{p^{\beta-\alpha}}.$$

By Hensel's lemma, there are  $1 + \left(\frac{\Delta'}{p}\right)$  solutions  $x_1 \pmod{p^{\beta-\alpha}}$  to this congruence, and so in total there are most  $2p^{\alpha/2}$  values of  $x_1$  modulo  $p^{\beta-\frac{\alpha}{2}}$ , giving (3.2).  $\square$



**3.2. The bounds on  $g(\chi, \psi)$ .** Recall that  $g(\chi, \psi)$  is defined by (1.5), and that both  $\chi$  and  $\psi$  are primitive characters modulo  $q = p^\beta$ . Anticipating some future simplifications, we apply the simple change of variables  $t \rightarrow t - 1$  and  $u \rightarrow u - 1$  giving

$$(3.3) \quad g(\chi, \psi) = \sum_{t, u \pmod{p^\beta}}^* \chi\left(\frac{u(t-1)}{t(u-1)}\right) \psi(ut - t - u),$$

where we recall that the asterisk on the sum means that the sum is restricted to  $u, t$  such that the denominator of  $\frac{u(t-1)}{t(u-1)}$  is coprime to  $p$ .

**Remark 3.2.** Note that if  $p = 2$  and  $q = p^\beta$ ,  $\beta \geq 1$ , then  $g(\chi, \psi)$  trivially vanishes, since  $t(t+1)$  is even for all  $t \in \mathbb{Z}$ .

**Theorem 3.3.** Suppose  $q = p^\beta$  with  $p$  odd and  $\beta = 2\alpha$ . Then

$$|g(\chi, \psi)| \leq q\rho(\Delta, p^\alpha),$$

where  $\Delta = A^2 + 4$  and  $A \equiv \ell_\chi \overline{\ell_\psi} \pmod{p^{\beta-1}}$ .

**Theorem 3.4.** Suppose  $q = p^\beta$  with  $p$  odd and  $\beta = 2\alpha + 1$ ,  $\alpha \geq 1$ . Then

$$(3.4) \quad |g(\chi, \psi)| \leq \begin{cases} 2q, & p \nmid \Delta, \\ qp^{1/2} \delta(p^2 | \Delta) \rho(p^{-2}\Delta, p^{\alpha-1}), & p | \Delta. \end{cases}$$

where  $\Delta = A^2 + 4$  with  $A \equiv \ell_\chi \overline{\ell_\psi} \pmod{p^{\beta-1}}$ .

*Proof of Theorem 3.3.* The sum (3.3) falls into the template of Lemma 2.2, with

$$(f_1(t, u), f_2(t, u)) = \left( \frac{u(t-1)}{t(u-1)}, tu - t - u \right), \quad (\chi_1, \chi_2) = (\chi, \psi).$$

No additive character is present, of course. A short calculation gives

$$(3.5) \quad (\log f)' = \begin{pmatrix} \frac{1}{\frac{t(t-1)}{u-1}} & \frac{-1}{\frac{u(u-1)}{t-1}} \\ \frac{t(t-1)}{ut-t-u} & \frac{t-1}{ut-t-u} \end{pmatrix}.$$

Note that the vanishing  $(\bmod p^\alpha)$  of the determinant of  $(\log f)'$  is equivalent to

$$(3.6) \quad u \equiv -t \pmod{p^\alpha},$$

and that

$$(3.7) \quad (\log f)'|_{u=-t} = \begin{pmatrix} \frac{1}{\frac{t(t-1)}{t+1}} & \frac{-1}{\frac{t(t+1)}{1-t}} \\ \frac{t(t-1)}{t^2} & \frac{t-1}{t^2} \end{pmatrix}.$$

By Lemma 2.7, the condition (2.5) is seen to be equivalent to (3.6) combined with

$$(3.8) \quad \frac{\ell_\chi}{t(t-1)} + \ell_\psi \frac{t+1}{t^2} \equiv 0 \pmod{p^\alpha}.$$

Simplifying (3.8), we obtain the equivalent congruence

$$(3.9) \quad t^2 + At - 1 \equiv 0 \pmod{p^\alpha}, \quad A \equiv \ell_\chi \overline{\ell_\psi} \pmod{p^\alpha}.$$

Hence,  $|g(\chi, \psi)| \leq q\rho(\Delta, p^\alpha)$ , as claimed.  $\square$

*Proof of Theorem 3.4.* The beginning steps of the proof are identical to those of Theorem 3.3; the linear congruences in both cases are the same, so we obtain that

$$g(\chi, \psi) = p^{2\alpha} \sum_{\substack{t_0, u_0 \pmod{p^\alpha} \\ (3.6) \text{ and } (3.9) \text{ hold}}}^* \chi(f_1(t_0, u_0))\psi(f_2(t_0, u_0))G_p(Q, L),$$

What is new is the presence of the quadratic Gauss sum  $G_p$ , so we next focus on this aspect. Note that the quadratic form  $Q$  present in  $G_p(Q, L)$  is given with respect to the standard basis by

$$(3.10) \quad 2Q = \ell_\chi(\log f_1)'' + \ell_\psi(\log f_2)''$$

Working in  $\mathbb{F}_p$  until further notice, the Hessian of  $\log f_1$  is

$$\begin{pmatrix} t^{-2} - (t-1)^{-2} & 0 \\ 0 & -u^{-2} + (u-1)^{-2} \end{pmatrix} = \begin{pmatrix} \frac{-2t+1}{t^2(t-1)^2} & 0 \\ 0 & \frac{-2t-1}{t^2(t+1)^2} \end{pmatrix},$$

by simplifying with (3.6). The Hessian of  $\log f_2$  is

$$\begin{pmatrix} -\frac{(u-1)^2}{(tu-t-u)^2} & \frac{-1}{(tu-t-u)^2} \\ \frac{-1}{(tu-t-u)^2} & -\frac{-(t-1)^2}{(tu-t-u)^2} \end{pmatrix} = \begin{pmatrix} \frac{-(t+1)^2}{t^4} & -t^{-4} \\ -t^{-4} & \frac{-(t-1)^2}{t^4} \end{pmatrix}.$$

Therefore,

$$(3.11) \quad 2Q = \frac{-\ell_\psi}{t^2} \left[ \begin{pmatrix} \frac{(t+1)^2}{t^2} & t^{-2} \\ t^{-2} & \frac{(t-1)^2}{t^2} \end{pmatrix} + \begin{pmatrix} \frac{A(2t-1)}{(t-1)^2} & 0 \\ 0 & \frac{A(2t+1)}{(t+1)^2} \end{pmatrix} \right].$$

Using a computer algebra package, we evaluate the determinant of the expression in square brackets above as

$$1 + \frac{4A}{t} - \frac{2}{t^2} + \frac{A^2(4t^2-1)}{(t^2-1)^2} = \frac{1}{t^2}(5t^2 + 4At - 3) = \frac{2-At}{t^2},$$

using  $t^2 - 1 = -At$ . Therefore, the determinant vanishes if and only if  $t = 2/A$ .

By Lemma 2.4, to determine the size of  $|G_p(Q, M)|$  we need the rank of  $Q$ . It is clear from (3.11) that  $Q$  does not have rank 0. Therefore,  $Q$  has rank 1 if the determinant vanishes, and rank 2 otherwise.

Next we note that the two algebraic equations  $t^2 + At - 1 = 0$  and  $t = 2/A$  have a common solution in  $\mathbb{F}_p$  if and only if  $A^2 + 4 = 0$  in  $\mathbb{F}_p$ , i.e.  $p|\Delta$ . Hence, if  $p \nmid \Delta$ , then  $|G_p(Q, M)| = p$ , and so  $|g(\chi, \psi)| \leq qp(\Delta, p^\alpha) \leq 2q$ , as desired. If  $p|\Delta$ , then  $Q$  has rank 1, so we obtain

$$|g(\chi, \psi)| \leq p^{2\alpha} \sum_{\substack{t_0 \pmod{p^\alpha} \\ t_0^2 + At_0 - 1 \equiv 0 \pmod{p^\alpha}}} p^{3/2} = qp^{1/2}\rho(\Delta, p^\alpha).$$

This bound is not as strong as (3.4); we will next gain some extra information by studying the behavior of the linear form  $L$  restricted to the isotropic subspace of  $Q$ . Note that  $\Delta = 0$  means  $A^2 = -4$ , whence  $t = -A/2 = 2/A$  and so  $t^2 = -1$ . Therefore  $(t+1)^2 = 2t$  and  $(t-1)^2 = -2t$ , and we can simplify (3.11) as

$$2Q = \ell_\psi \left[ \begin{pmatrix} -2t & -1 \\ -1 & 2t \end{pmatrix} + \begin{pmatrix} \frac{A(2t-1)}{-2t} & 0 \\ 0 & \frac{A(2t+1)}{2t} \end{pmatrix} \right] = \ell_\psi \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Hence the isotropic subspace of  $Q$  is spanned by the vector  $(1, -1)^\top$ .

Next we work out an easily-checked characterization for the linear form  $L$  to be trivial on this isotropic subspace. By (2.8), (3.7), and the above calculation of the isotropic subspace, this means

$$p^{-\alpha} \left( \frac{\ell_\chi}{t-1} + \frac{\ell_\psi(1+t)}{t} \right) - p^{-\alpha} \left( \frac{-\ell_\chi}{t+1} + \frac{\ell_\psi(1-t)}{t} \right) \equiv 0 \pmod{p},$$

which reduces to  $t$  satisfying

$$t^2 + At - 1 \equiv 0 \pmod{p^{\alpha+1}}.$$

Thus the number of  $t_0$  to be estimated is

$$(3.12) \quad \#\{t_0 \pmod{p^\alpha} : (t_0 + A/2)^2 \equiv \Delta \pmod{p^{\alpha+1}}\},$$

and we study this a bit more closely (which along the way will confirm this quantity is well-defined). This count equals  $\#\{x \pmod{p^\alpha} : x^2 \equiv \Delta \pmod{p^{\alpha+1}}\}$ . Since  $p|\Delta$ , then  $p|x$  also, so (3.12) equals  $\#\{x_1 \pmod{p^{\alpha-1}} : x_1^2 \equiv \frac{\Delta}{p^2} \pmod{p^{\alpha-1}}\}$ , which is well-defined. Therefore, we obtain a more refined bound

$$|g(\chi, \psi)| \leq qp^{1/2} \delta(p^2|\Delta) \rho(p^{-2}\Delta, p^{\alpha-1}). \quad \square$$

#### 4. BOUNDING THE CUBIC MOMENT

In this section, we prove [PY2, Conj. 8.2] assuming Theorem 1.4. Conjecture 8.2 of [PY2] implies the cubic moment bounds (Theorems 1.2 and 1.3) and hence the Weyl bound (Theorem 1.1). The precise statement of [PY2, Conj. 8.2] appears as Lemma 4.2, below. The proof of Theorem 1.4 is deferred to Sections 8 and 9.

We begin by reviewing the notation and re-stating this conjecture. We have a Dirichlet series

$$(4.1) \quad Z(s_1, s_2, s_3, s_4) = \frac{1}{\varphi(q)} \sum_{\psi \pmod{q}} \frac{L(s_1, \psi) L(s_2, \psi) L(s_3, \psi) L(s_4, \bar{\psi})}{\zeta^{(q)}(s_1 + s_4)} Z_{\text{fin}},$$

where  $Z_{\text{fin}} = Z_{\text{fin}}(\chi, \psi)$ ,  $|Z_{\text{fin}}| = |\prod_{p|q} Z_{\text{fin},p}|$ , and  $Z_{\text{fin},p}$  is a certain Dirichlet series supported on powers of  $p$ . Its precise definition is not necessary here, but rather we quote Lemma 7.1 from [PY2].

**Lemma 4.1.** *Let  $q = p^\beta$ , and let  $\chi = \chi_p$  be primitive modulo  $q$ . The series  $Z_{\text{fin},p}$  converges absolutely when  $\text{Re}(s_j) = \sigma_j > 0$  for all  $j = 1, 2, 3, 4$ . If  $\sigma_j \geq \sigma > 1/2$  for all  $j$ , then*

$$(4.2) \quad Z_{\text{fin},p}(s_1, s_2, s_3, s_4) \ll_{\sigma} \delta_{\psi} q^{1/2} |g(\chi, \psi)| + q^{3/2+\varepsilon},$$

where  $\delta_{\psi} = 1$  if  $\psi$  is primitive, and 0 otherwise. If  $\sigma_j \geq \sigma > 1$  for all  $j$ , and  $\psi_p$  is the trivial character, then

$$(4.3) \quad Z_{\text{fin},p}(s_1, s_2, s_3, s_4) \ll_{\sigma} q^{1+\varepsilon}.$$

We remark that Lemma 4.1 appeared as [PY2, Lem. 7.1], however there it was conditional on [PY2, Conj. 6.6] which has been proved here as Lemma 2.8.

As in [PY2], it is helpful to treat the trivial character separately. To this end, write  $Z = Z_0 + Z_1$ , where  $Z_0$  is the contribution to  $Z$  from the trivial character. Now we state the main lemma.

**Lemma 4.2.** *The functions  $Z_0$  and  $Z_1$  satisfy the following properties. Firstly,  $Z_0$  is meromorphic for  $\operatorname{Re}(s_j) \geq \sigma > 1/2$  for all  $j$  and analytic for  $\operatorname{Re}(s_j) \geq \sigma > 1$  for all  $j$ . In this domain, it may only have polar divisors on the hyperplanes  $s_j = 1$ . In the region  $\operatorname{Re}(s_j) \geq \sigma > 1$  it satisfies the bound*

$$(4.4) \quad Z_0(s_1, s_2, s_3, s_4) \ll_{\sigma} q^{\varepsilon}.$$

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

$$(4.5) \quad \int_{-T}^T |Z_1(\sigma + it, \sigma + it, \sigma + it, \sigma - it)| dt \ll q^{3/2+\varepsilon} T^{1+\varepsilon},$$

*for  $T \geq 1$ . The same bounds stated for  $Z_1$  also hold for  $Z_0$  (in an even stronger form), provided  $1/2 \leq \operatorname{Re}(s_j) \leq 0.99$ .*

*Remark.* Theorem 1.4 is the crucial new ingredient in the proof of Lemma 4.2 in the case that  $q$  is not cube-free.

*Proof.* The holomorphic (resp. meromorphic) continuation of  $Z_1$  (resp.  $Z_0$ ) follows from the definition (4.1) and Lemma 4.1. All the required properties of  $Z_0$  follow from Lemma 4.1, so we now focus on  $Z_1$ .

Supposing that  $\operatorname{Re}(s_j) > 1/2$  for  $j = 1, 2, 3, 4$ , we have

$$(4.6) \quad Z_1(s_1, s_2, s_3, s_4) \ll \frac{(1+|t|)^{\varepsilon}}{q^{1-\varepsilon}} \sum_{\psi \neq \psi_0} |L(s_1, \psi)L(s_2, \psi)L(s_3, \psi)L(s_4, \bar{\psi})| \prod_{p|q} |Z_{\text{fin},p}|.$$

From Lemma 4.1 we have for  $\chi_p, \psi_p$  the  $p$ -parts of  $\chi, \psi$  respectively, each modulo  $p^{\beta}$ ,

$$|Z_{\text{fin},p}| \ll p^{(\frac{3}{2}+\varepsilon)\beta} \left( \delta_{\psi_p} \frac{|g(\chi_p, \psi_p)|}{p^{\beta}} + 1 \right).$$

Recall from Remark 3.2 that  $g(\chi_p, \psi_p) = 0$  if  $p = 2$ , so for the forthcoming analysis of  $Z_{\text{fin},p}$  we largely assume  $p$  is odd. In [PY2, Thm. 6.9] it was shown that if  $\beta = 1$  then  $|g(\chi_p, \psi_p)| \leq Cp$  for some absolute constant  $C \geq 2$ . On the other hand, when  $\beta \geq 2$ , we see from Theorems 3.3 and 3.4 that  $|Z_{\text{fin},p}|$  is controlled by the quantity  $\Delta_p = A^2 + 4 = (\ell_{\chi} \bar{\ell}_{\psi})^2 + 4 \pmod{p^{\beta-1}}$  where  $p^{\beta} \parallel q$ . Therefore it is natural to parametrize the sum in (4.6) over the possible values of the parameters  $\Delta_p$ . To this end, for  $\beta - 1 \geq \alpha \geq 0$  and  $C$  the above absolute constant, let

$$m(\alpha, \beta) = \inf \left\{ m \in \frac{1}{2}\mathbb{Z} : \max_{\substack{\psi_p \pmod{p^{\beta}} \text{ prim.} \\ v_p(\Delta_p) = \alpha}} \frac{|g(\chi_p, \psi_p)|}{p^{\beta}} \leq Cp^m \right\},$$

which depends on  $p$  and  $\chi_p$ , but we suppress this from the notation. For  $p = 2$ ,  $m(\alpha, \beta) = -\infty$ . For  $a \mid \frac{q}{\tilde{q}}$  with  $\tilde{q} = \prod_{p|q} p$ , let

$$M(a, q) = \prod_{p^{\beta} \parallel q} p^{m(\alpha, \beta)}, \quad \text{where } \alpha = v_p(a).$$

Write  $\Delta = \Delta(\psi) \in [1, \frac{q}{\tilde{q}}] \subset \mathbb{Z}$  with  $\Delta \equiv \Delta_p \pmod{p^{\beta-1}}$  for each  $p \mid q$ . Note the condition that  $v_p(\Delta_p) = \alpha$  for all  $p \mid q$  is equivalent to  $a \parallel \Delta$ . Then we have

$$(4.7) \quad Z_1(s_1, s_2, s_3, s_4) \ll q^{\frac{1}{2}+\varepsilon} (1+|t|)^{\varepsilon} \sum_{a \mid \frac{q}{\tilde{q}}} M(a, q) \sum_{\psi: \Delta(\psi) \equiv 0 \pmod{a}} |L(s_1, \psi)L(s_2, \psi)L(s_3, \psi)L(s_4, \bar{\psi})|,$$

where we over-extended the condition  $a \parallel \Delta$  to  $\Delta \equiv 0 \pmod{a}$ . With an eye towards applying Theorem 1.4, we next break up (4.7) over cosets. Let  $G = \{\psi \pmod{q}\}$ , and  $H_a$  be the subgroup  $H_a = \{\psi \pmod{q/a}\}$ .

Lemma 2.1 implies that  $\psi, \psi' \in G$  are in the same  $H_a$ -coset if and only if  $\ell_\psi \equiv \ell_{\psi'} \pmod{p^{v_p(a)}}$  for each  $p|q$ . Hence if  $\psi, \psi'$  are in the same  $H_a$ -coset, then  $\Delta(\psi) \equiv \Delta(\psi') \pmod{a}$ . Thus

$$\begin{aligned} & Z_1(s_1, s_2, s_3, s_4) \\ & \ll q^{\frac{1}{2}+\varepsilon} (1+|t|)^\varepsilon \sum_{a|\frac{q}{4}} M(a, q) \sum_{\substack{\theta \in G/H_a \\ \Delta(\theta) \equiv 0 \pmod{a}}} \sum_{\eta \in H_a} |L(s_1, \eta, \theta) L(s_2, \eta, \theta) L(s_3, \eta, \theta) L(s_4, \overline{\eta, \theta})|. \end{aligned}$$

Next, we introduce an integral as in (4.5), and apply Theorem 1.4 to find

$$(4.8) \quad \frac{1}{T^{1+\varepsilon}} \int_{-T}^T |Z_1| dt \ll q^{\frac{1}{2}+\varepsilon} \sum_{a|\frac{q}{4}} M(a, q) \sum_{\substack{\theta \in G/H_a \\ \Delta(\theta) \equiv 0 \pmod{a}}} \text{lcm}(q/a, q^*),$$

where  $Z_1$  is shorthand for  $Z_1(\sigma + it, \sigma + it, \sigma + it, \sigma - it)$ .

The right hand side of (4.8) is a multiplicative function of  $q$ , and so is the desired bound of  $q^{3/2+\varepsilon}$ , so it suffices to work with  $q = p^\beta$  an odd prime power, which we henceforth assume. Note that there are at most two  $\theta \in G/H_{p^\alpha}$  satisfying the condition  $\Delta \equiv 0 \pmod{p^\alpha}$ . Indeed,  $\Delta \equiv 0 \pmod{p^\alpha}$  means that  $\ell_\chi^2 = -4\ell_\psi^2 \pmod{p^\alpha}$ , which has at most two solutions  $\ell_\psi \pmod{p^\alpha}$ , since  $(\ell_\chi, p) = 1$ . Thus the right hand side of (4.8) takes the form

$$(4.9) \quad p^{(\frac{1}{2}+\varepsilon)\beta} \sum_{\alpha=0}^{\beta-1} p^{m(\alpha, \beta) + \max(\beta - \alpha, \lceil \frac{2\beta}{3} \rceil)}.$$

To finish the proof of the lemma, it suffices to show for all  $0 \leq \alpha \leq \beta - 1$  the inequality

$$(4.10) \quad m(\alpha, \beta) + \max(\beta - \alpha, \lceil \frac{2\beta}{3} \rceil) \leq \beta.$$

By [PY2, Thm. 6.9], and Theorems 3.3 and 3.4, we have

$$\frac{|g(\chi_p, \psi_p)|}{p^\beta} \leq \begin{cases} \rho(\Delta, p^{\beta/2}) & \text{for } \beta \text{ even,} \\ C & \text{for } \beta = 1, \\ p^{1/2} \rho(\frac{\Delta}{p^2}, p^{\frac{\beta-3}{2}}) & \text{for } \beta \text{ odd, } \beta \geq 3, p^2 | \Delta, \\ 2 & \text{for } \beta \text{ odd, } \beta \geq 3, p^2 \nmid \Delta. \end{cases}$$

By Lemma 3.1, we get for  $\beta$  even

$$(4.11) \quad m(\alpha, \beta) \leq \begin{cases} -\infty & \text{for } \alpha \text{ odd, } \alpha < \beta/2, \\ \alpha/2 & \text{for } \alpha \text{ even, } \alpha < \beta/2, \\ \lfloor \beta/4 \rfloor & \text{for } \alpha \geq \beta/2, \end{cases}$$

and for  $\beta$  odd,

$$m(\alpha, \beta) \leq \begin{cases} 0 & \text{for } \alpha = 0, \\ -\infty & \text{for } \alpha \text{ odd, } \alpha < \frac{\beta+1}{2}, \\ \frac{\alpha-1}{2} & \text{for } \alpha \text{ even, } 2 \leq \alpha < \frac{\beta+1}{2}, \\ \lfloor \frac{\beta+1}{4} \rfloor - \frac{1}{2} & \text{for } \alpha \geq \frac{\beta+1}{2}. \end{cases}$$

We proceed to prove (4.10). First suppose  $\beta$  is even, so  $m(\alpha, \beta)$  is bounded by (4.11). If  $\alpha \geq \beta/2$ , then  $\max(\beta - \alpha, \lceil 2\beta/3 \rceil) = \lceil 2\beta/3 \rceil$ , and it reduces to checking  $\lfloor \beta/4 \rfloor + \lceil 2\beta/3 \rceil \leq \beta$ . To show this last inequality, it suffices to check it for each  $\beta \in \{0, 2, 4, 6, 8, 10\}$ , which may as well be done by brute force using a computer. For  $\alpha < \beta/2$ , we have  $\alpha/2 + (\beta - \alpha) \leq \beta$ , as well as  $\alpha/2 + \lceil 2\beta/3 \rceil \leq \lfloor \beta/4 \rfloor + \lceil 2\beta/3 \rceil \leq \beta$ , so we are done. Similarly easy arguments hold when  $\beta$  is odd, and we omit the details.  $\square$

## 5. REDUCTION OF THEOREM 1.4 TO THEOREM 1.5

In this section, we prove Theorem 1.4, subject to the veracity of Theorem 1.5. The rest of the paper is then devoted to the proof of Theorem 1.5.

First note that by positivity, to prove Theorem 1.4, it suffices to consider the case  $q^*|d$ , which means  $q^2|d^3$ . By an approximate functional equation, dyadic partition of unity, and Cauchy's inequality applied on the dyadic sum, it suffices to show

$$\mathcal{M}(N, d, q, T) := \int_{-\infty}^{\infty} w_0\left(\frac{t}{T}\right) \sum_{\psi \pmod{d}} \frac{d}{\varphi(d)} \left| \sum_n w_N(n) \tau(n) \psi(n) \chi(n) n^{-it} \right|^2 dt \ll NdT(qT)^\varepsilon,$$

where  $w_N(x)$  is a smooth function supported on  $[N, 2N]$ , satisfying  $w_N^{(j)}(x) \ll x^{-j}$ , for all  $j \geq 0$ , and  $w_0$  is a fixed smooth nonnegative function. Moreover, we may assume

$$(5.1) \quad N \ll (qT)^{1+\varepsilon}.$$

Opening the square and executing the  $\psi$  sum and  $t$ -integral, we have

$$\mathcal{M}(N, d, q, T) = dT \sum_{m \equiv n \pmod{d}} \tau(m) \chi(m) \tau(n) \bar{\chi}(n) \widehat{w}_0\left(\frac{T}{2\pi} \log\left(\frac{m}{n}\right)\right) w_N(m) w_N(n),$$

where  $\widehat{w}_0(y) = \int_{-\infty}^{\infty} w_0(t) e(-ty) dt$  is the standard Fourier transform. The contribution from the diagonal terms  $m = n$  give a main term of size  $O(NdTq^\varepsilon)$ , which is acceptable.

Next consider the off-diagonal terms. By symmetry, it suffices to consider the terms with  $m > n$ , in which case we write  $m = n + h$ , with  $h \geq 1$ , and  $d|h$ . By the rapid decay of  $\widehat{w}_0$ , the sum over  $h$  may be truncated at  $h \ll H$  where  $H = \frac{N}{T}(Nq)^\varepsilon$ . By the positivity in (1.8), we may also assume  $T \gg (qN)^\varepsilon$  so that  $H \ll N$ . We also open  $\tau(n) = \sum_{n_1 n_2 = n} 1$  and employ dyadic partitions of unity to the sums over  $n_1$  and  $n_2$ . Let  $\mathcal{M}_1(N_1, N_2, d, q, T)$  denote the contribution of these terms to  $\mathcal{M}(N, d, q, T)$ , where  $N_1 N_2 \asymp N$  and  $n_j \asymp N_j$  for  $j = 1, 2$ . Then

$$(dT)^{-1} \mathcal{M}_1(N, d, q, T) = S(\chi),$$

where the weight function  $w(n_1, n_2, h)$  is given by

$$w(x, y, z) = \widehat{w}_0\left(\frac{T}{2\pi} \log\left(\frac{xy + z}{xy}\right)\right) w_N(xy + z) w_N(xy) \omega_1(x) \omega_2(y),$$

where  $\omega_1$  and  $\omega_2$  are part of the dyadic partitions of unity. It is easy to check that  $w(x, y, z)$  satisfies (1.13).

Theorem 1.5 will complete the proof of Theorem 1.4, since  $\frac{H}{q} \ll \frac{N}{qT}(qN)^\varepsilon \ll (qT)^\varepsilon$ .

## 6. AUTOMORPHIC FORMS

**6.1. Fourier expansion.** In this section we recall the Fourier expansions of automorphic forms on  $\mathrm{GL}_2$  over  $\mathbb{Q}$ . Using canonical inner products on Whittaker models as in [MV], we obtain particularly pleasant normalizations of Fourier expansions and Bruggeman-Kuznetsov formulas. To discuss these, we work in greater generality than is strictly required for the other sections of this paper.

Let  $(\pi, V)$  be a standard generic automorphic representation of  $\mathrm{GL}_2/\mathbb{Q}$  of conductor  $\mathfrak{c}(\pi)$  and analytic conductor  $C(\pi)$  (for a definition, see [MV, 3.1.8]). By “standard” here we mean, following [MV, 2.2.1], that  $\pi$  occurs in the spectral decomposition of the space of automorphic forms. In particular, it is abstractly unitarizable. For  $q \in \mathbb{N}$ , let

$$K_1(q) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\widehat{\mathbb{Z}}) : c \in (q), \quad d \in 1 + (q) \right\} \subset \mathrm{GL}_2(\mathbb{A}).$$

The subspace of  $V$  consisting of right  $K_1(\mathfrak{c}(\pi))$ -invariant vectors of minimal non-negative  $\mathrm{SO}_2(\mathbb{R})$ -weight is 1-dimensional (see [C], [De1, §2.2]). Any  $\phi$  belonging to this 1-dimensional subspace is called a *newvector*. In Theorem 6.1, we make an explicit choice of a distinguished newvector in this 1-dimensional space using canonical inner products on Whittakers models.

To give the precise statement, we must set up some notation. Fix  $\psi : \mathbb{A}/\mathbb{Q} \rightarrow \mathbb{C}^\times$  the unique additive character which coincides with  $e(x)$  on  $\mathbb{R}$ . For any place  $v$  of  $\mathbb{Q}$ , let us denote by  $\psi_v$  the restriction of  $\psi$  to  $\mathbb{Q}_v$ . Let  $\mathbb{X} = \mathrm{PGL}_2(\mathbb{Q}) \backslash \mathrm{PGL}_2(\mathbb{A})$ . The space  $\mathbb{X}$  has finite measure, which we normalize to be probability measure. Warning: Michel-Venkatesh use the push-forward measure on  $\mathbb{X}$ , under which it has volume  $2\xi(2) = \pi/3$ , see [MV, 4.1.2].

Given  $\pi = \bigotimes_v \pi_v$  unitary, for  $v = p < \infty$  let  $\phi_p : \mathbb{Q}_p^\times \rightarrow \mathbb{C}$  be a local newvector for  $\pi_p$  in the Kirillov model  $\mathcal{W}(\pi_p, \psi_p)$ , normalized so that  $\phi_p(1) = 1$ . Explicit formulas for  $\phi_p$  are well-known, see e.g. [Sch, §2.4 Summary] for a nice presentation. In particular,  $\mathrm{supp}(\phi_p) \subseteq \mathbb{Z}_p$ ,  $\phi_p(x)$  only depends on  $|x|_p$ , and  $\lambda_\pi(n) = |n|^{1/2} \prod_p \phi_p(p^{v_p(n)})$  coincides with the  $n$ th Hecke eigenvalue of  $\pi$  normalized so that the Ramanujan conjecture predicts that  $|\lambda_\pi(p)| \leq 2$ . We also have  $|\phi_p(x)| \leq |x|^{1/2}$  for all  $x \in \mathbb{Q}_p^\times$  if  $\pi_p$  is ramified, and in particular,

$$(6.1) \quad |\lambda_\pi(p)| \leq 1 \quad \text{if} \quad p \mid \mathfrak{c}(\pi).$$

Following the notation in [MV, 4.1.5], for  $L$  a meromorphic function, we write  $L^*(s_0)$  for the leading coefficient in the Laurent series of  $L(s)$  at  $s = s_0$ . By the analytic continuation of Rankin-Selberg  $L$ -functions, the series

$$(6.2) \quad \mathcal{L}_\pi(s) := \sum_{n \geq 1} \frac{|\lambda_\pi(n)|^2}{n^s}$$

admits a meromorphic continuation to  $\mathrm{Re}(s) > 1/2$ , with no poles except at  $s = 1$ . We have [Iw2, HL]

$$(6.3) \quad \mathcal{L}_\pi^*(1) = C(\pi)^{o(1)}.$$

Lastly, for  $\phi \in \pi$  recall from [MV, §2.2.2] the canonical norm  $\|\phi\|_{can}$  on the space of  $\pi$ . (Note that there is a missing factor of  $\xi_F^*(1)/\xi_F(2)$  on the right hand side of [MV, (2.3)]. That this factor is missing is suggested by the notational conventions for infinite products in section 4.1.5; also compare (2.3) to e.g. (4.16) or (4.28) for a precise check.)

Let  $W_{\lambda, \mu}(z)$  be the Whittaker function defined and normalized as in [GR, 9.220.4] and  $K_\nu$  be the standard  $K$ -Bessel function, as in [GR, 9.235.2]. Let  $\alpha$  be the character of  $\mathbb{R}^\times$  given by  $x \mapsto |x|$  and  $\mathrm{sgn}$  be the character of  $\mathbb{R}^\times$  given by  $x \mapsto x/|x|$ .

**Theorem 6.1.** *Let  $\pi = \bigotimes_v \pi_v$  be a standard generic automorphic representation of  $\mathrm{GL}_2/\mathbb{Q}$  with finite-order central character.*

- (1) *For  $\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R}) \hookrightarrow \mathrm{GL}_2(\mathbb{A})$ , a newvector  $\phi$  for  $\pi$  admits a Fourier expansion of the form*

$$\phi\left(\begin{pmatrix} y & x \\ & 1 \end{pmatrix}\right) = c_\phi(y) + \sum_{n \neq 0} \frac{\rho_\phi(n)}{|n|^{1/2}} W(ny) e(nx),$$

where  $c_\phi(y)$  is a (possibly vanishing) constant term, the coefficients  $\rho_\phi(n) = \rho_\phi(1)\lambda_\pi(|n|)$ , the function  $W$  is a minimal non-negative weight vector in the Kirillov model  $\mathcal{W} = \mathcal{W}(\pi_\infty, \psi_\infty)$  with  $\|W\|_{L^2(\mathcal{W})}^2 = 1$ , and

$$(6.4) \quad \|\phi\|_{can}^2 = 2\xi(2)|\rho_\phi(1)|^2 \mathcal{L}_\pi^*(1).$$

- (2) *A Whittaker function  $W$  satisfying the hypotheses of the previous point can be given explicitly as follows.*

- (a) *If  $\pi_\infty \simeq \pi(\alpha^s \mathrm{sgn}^\epsilon, \alpha^{-s} \mathrm{sgn}^\epsilon)$  with  $\epsilon \in \{0, 1\}$  and  $s \in i\mathbb{R} \cup (-1/2, 1/2)$ , then we have*

$$W(y) = (\mathrm{sgn} y)^\epsilon \left(\frac{\cos \pi s}{\pi}\right)^{1/2} W_{0,s}(4\pi|y|) = (\mathrm{sgn} y)^\epsilon \left(\frac{\cos \pi s}{\pi}\right)^{1/2} 2\sqrt{|y|} K_s(2\pi|y|).$$

- (b) *If  $\pi_\infty \simeq \pi(\alpha^s \mathrm{sgn}, \alpha^{-s})$  with  $s = it \in i\mathbb{R}$ , then we have*

$$W(y) = \left(\frac{\sinh \pi t}{\pi t \mathrm{sgn} y}\right)^{1/2} W_{\frac{1}{2} \mathrm{sgn} y, it}(4\pi|y|).$$

- (c) *If  $\pi_\infty \simeq \pi(\alpha^s \mathrm{sgn}, \alpha^{-s})$  with  $s \in (-1/2, 0) \cup (0, 1/2)$ , then we have*

$$W(y) = \left(\frac{\cos \frac{\pi s}{2}}{\left(\frac{s}{2}\right) \mathrm{sgn} y - 1}\right)^{1/2} W_{\frac{1}{2} \mathrm{sgn} y, \frac{s}{2}}(4\pi|y|).$$

- (d) *If  $\pi_\infty \simeq \sigma(\chi_1, \chi_2)$  with  $\chi_1 \chi_2^{-1} = \alpha^s \mathrm{sgn}^m$ ,  $m \in \{0, 1\}$  and  $s - m \in 1 + 2\mathbb{Z}_{\geq 0}$  or  $(s, m) = (0, 1)$ , then writing  $k = s + 1$  we have*

$$W(y) = \Gamma(k)^{-1/2} W_{\frac{k}{2}, \frac{k-1}{2}}(4\pi y) \delta(y > 0) = \left(\frac{(4\pi y)^k}{\Gamma(k)}\right)^{1/2} e^{-2\pi y} \delta(y > 0).$$

*There is a unique newvector  $\phi$  in  $\pi$  with  $W$  as above,  $\|\phi\|_{can}^2 = 1$ , and  $\rho_\phi(1) > 0$ .*

Remark 1: The Selberg eigenvalue conjecture predicts that the cases (2a) and (2c) above with  $s$  real and non-zero never occur as local components of any automorphic representation, but one cannot at present rule out this possibility.

Remark 2: The explicit choice of  $W$  given in part (2) of the above theorem is used later to justify the choice of normalization in Theorem 6.9 (see the remarks following it). In addition, we believe it could be valuable to record Theorem 6.1 for the sake of reference.

Remark 3: In addition to  $\|\phi\|_{can}$ , Michel-Venkatesh define an Eisenstein norm  $\|\phi\|_{\mathrm{Eis}}$  on the space of  $\pi$ , see [MV, §2.2.1]. Correcting [MV, (2.3)] as pointed out following (6.3), for a number field  $F$  and  $\phi \in \pi$ , we have  $\|\phi\|_{can}^2 = 2\xi_F^*(1)\|\phi\|_{\mathrm{Eis}}^2$  if  $\pi$  is Eisenstein and non-singular, and  $\|\phi\|_{can}^2 = \|\phi\|_{L^2(\mathbb{X}, \mathrm{push})}^2$  if  $\pi$  is cuspidal and  $\mathbb{X}$  is given the push-forward measure (cf. [MV, Lem. 2.2.3]).

Let  $S_*(q, \chi)$  denote either  $S_{it_j}(q, \chi)$  or  $S_k(q, \chi)$ , the vector space of Maass (resp. holomorphic) cusp forms of level  $q$ , central character  $\chi$  and spectral parameter  $t_j$  (resp. weight  $k$ ).



There is a natural embedding  $f \mapsto \phi_f$  of  $S_*(q, \chi)$  in the space of automorphic forms. We have in particular for  $f \in S_*(q, \chi)$  that

$$(6.5) \quad y^{\frac{k}{2}} f(x + iy) = \phi_f \left( \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \right)$$

and  $\langle f, f \rangle = \|\phi_f\|_{L^2(\mathbb{X})}^2$ , where the Petersson inner product is defined with respect to probability measure on  $\Gamma_0(q) \backslash \mathcal{H}$  (for details, see e.g. [KL1, Prop. 12.5]). We continue to write  $S_*(q, \chi)$  for the image of this space under the map  $f \mapsto \phi_f$  despite the abuse of notation.

Let  $\mathcal{E}_{it}(q, \chi)$  denote the vector space of Eisenstein series of level  $q$ , central character  $\chi$  and spectral parameter  $t$ , as in [Y2, §8.1]. As with cusp forms, the space  $\mathcal{E}_{it}(q, \chi)$  embeds in the space of automorphic forms by  $E \mapsto \phi_E$ . The second author defined a formal inner product  $\langle \cdot, \cdot \rangle_{\text{formal}}$  on  $\mathcal{E}_{it}(q, \chi)$  in loc. cit. by setting  $\frac{1}{4\pi} \langle E_{\mathfrak{a}}, E_{\mathfrak{b}} \rangle_{\text{formal}} = \delta_{\mathfrak{a}\mathfrak{b}}$  for Eisenstein series attached to singular cusps  $\mathfrak{a}, \mathfrak{b}$ , and extending linearly. By chasing definitions in [KL2, §5] and [Y2, (3.3), Lem. 8.3], one finds that  $\frac{1}{4\pi} \langle E, E \rangle_{\text{formal}} / \nu(q) = \|\phi_E\|_{\text{Eis}}^2$ , where  $\nu(q) = [\text{SL}_2(\mathbb{Z}) : \Gamma_0(q)]$ .

If  $f$  is an even (resp. odd) weight 0 Maass form or Eisenstein series  $f$  of spectral parameter  $t$ , then Theorem 6.1(2a) applies to  $\phi_f$  with  $s = 2it$  and  $\epsilon = 0$  (resp. 1). If  $f$  is a weight  $k$  holomorphic cusp form, then Theorem 6.1(2d) applies to  $\phi_f$ . Theorem 6.1(2b) and (2c) pertain to weight 1 Maass forms.

*Proof sketch.* If  $\pi$  is generic, then  $\phi \in \pi$  admits a Whittaker-Fourier expansion

$$\phi(g) = \phi_N(g) + \sum_{\alpha \in \mathbb{Q}^\times} W \left( \begin{pmatrix} \alpha & \\ & 1 \end{pmatrix} g \right),$$

where  $\phi_N(g) = \int_{\mathbb{A}/\mathbb{Q}} \phi(n(x)g) dx$  and  $W$  is a global Whittaker function. By expressing  $W$  in terms of local Whittaker functions  $W_v$  and restricting to  $g = \begin{pmatrix} y & x \\ & 1 \end{pmatrix} \times 1_{\text{fin}}$  we derive the Fourier expansion found in part (1) of Theorem 6.1. From the product over  $v < \infty$  of  $W_v$  one extracts the Hecke eigenvalue  $\lambda_\pi$ . Following [MV, §2.2.2], the canonical norm  $\|\phi\|_{\text{can}}^2$  is given by a regularized infinite product of local norms on Whittaker models, which leads to the relation (6.4).

If  $\pi_\infty$  is a unitary principal series, the formulas for the Whittaker function  $W_\infty$  in part (2) of the Theorem can be derived from the explicit isometry between the induced model and the Whittaker model given by [MV, (3.10)] and the integral formula [GR, 3.384.9] for  $W_{\lambda, \mu}(z)$ . If  $\pi_\infty$  is complementary series or discrete series, then [MV, (3.10)] still intertwines the induced model and Whittaker model but may no longer be an isometry. In these cases, we may compute  $\|W_\infty\|_{L^2(\mathcal{W})}^2$  by hand using [GR, 7.611.4] for complementary series and using the definition of  $\Gamma(s)$  for discrete series.  $\square$

**6.2. Twisting.** Let  $F$  be a non-archimedean local field, and  $\pi$  an irreducible, admissible, generic representation of  $\text{GL}_2(F)$  with central character  $\omega_\pi$ . Writing  $c$  for the conductor exponent, we say that  $\pi$  is *twist-minimal* if  $c(\pi) \leq c(\pi \otimes \chi)$  for all quasi-characters  $\chi$  of  $F^\times$ . The following lemma appears in e.g. [BLS, Lem. 1.4] or [CS, Lem. 2.7], and relies principally on [T, Prop. 3.4].

**Lemma 6.2.** *For all quasi-characters  $\chi$  of  $F^\times$  we have*

$$c(\pi \otimes \chi) \leq \max(c(\pi), c(\chi) + c(\omega_\pi \chi)),$$

*with equality if  $\pi$  is twist-minimal.*

If  $\pi$  is a global automorphic representation of  $\mathrm{GL}_2$ , then we say that  $\pi$  is twist-minimal at  $p$  if the associated local representation is twist-minimal. We say that  $\pi$  is (globally) twist-minimal if it is twist-minimal at all primes dividing its conductor. In that case, we have

$$(6.6) \quad \mathbf{c}(\pi \otimes \chi) = [\mathbf{c}(\pi), \mathbf{c}(\chi) \mathbf{c}(\omega_\pi \chi)],$$

where  $[m, n]$  denotes  $\mathrm{lcm}(m, n)$ .

**6.3. Cusps.** Our presentation of cusps and scaling matrices in this subsection is inspired by [NPS, §3.4.1]. Here we restrict our attention to cusps with respect to Hecke congruence subgroups  $\Gamma_0(q)$ . For more general co-compact subgroups of  $\mathrm{GL}_2^+(\mathbb{R})$ , see loc. cit.

The group  $\Gamma = \mathrm{SL}_2(\mathbb{Z})$  acts transitively on  $\mathbb{P}^1(\mathbb{Q})$  by fractional linear transformations. Let  $\Gamma_\infty = \{\pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z}\}$  be the stabilizer of  $\infty$  in  $\Gamma$ . Thus we may identify

$$(6.7) \quad \mathbb{P}^1(\mathbb{Q}) \simeq \Gamma/\Gamma_\infty$$

upon picking the base point of  $\mathbb{P}^1(\mathbb{Q})$  to be  $\infty$ .

**Definition 6.3.** *The set of left  $\Gamma_0(q)$ -orbits*

$$\mathcal{C}(q) := \Gamma_0(q) \backslash \Gamma/\Gamma_\infty$$

*is called the set of cusps of  $\Gamma_0(q)$ . A cusp  $\mathfrak{a}$  may be identified with a  $\Gamma_0(q)$ -orbit in  $\mathbb{P}^1(\mathbb{Q})$  via the bijection (6.7). The index  $w_{\mathfrak{a}} := [\mathrm{Stab}_\Gamma(\mathfrak{a}) : \mathrm{Stab}_{\Gamma_0(q)}(\mathfrak{a})]$  of a cusp  $\mathfrak{a} \in \mathcal{C}(q)$  is called the width of  $\mathfrak{a}$ .*

The notion of width of a cusp in Definition 6.3 matches the usual geometric intuition: choosing a fundamental domain  $\mathcal{F}_q$  for  $\Gamma_0(q) \backslash \mathcal{H}$  to be a union of translates of the standard fundamental domain  $\mathcal{F}$  for  $\Gamma \backslash \mathcal{H}$ , the width  $w_{\mathfrak{a}}$  is the number of translates of  $\mathcal{F}$  that touch  $\mathfrak{a}$  in  $\mathcal{F}_q$ . Another description of the width  $w_{\mathfrak{a}}$  is that  $w_{\mathfrak{a}} = [\Gamma_\infty : \Gamma_\infty \cap \tau^{-1} \Gamma_0(q) \tau]$ , where  $\tau \in \Gamma$  is any representative of  $\mathfrak{a}$ .

**Definition 6.4.** *If  $\tau \in \Gamma$  represents the cusp  $\mathfrak{a} \in \mathcal{C}(q)$ , then*

$$\sigma_{\mathfrak{a}} = \tau \begin{pmatrix} w_{\mathfrak{a}} & \\ & 1 \end{pmatrix}$$

*is called a scaling matrix for  $\mathfrak{a}$ .*

A scaling matrix for  $\mathfrak{a}$  satisfies  $\sigma_{\mathfrak{a}} \infty = \mathfrak{a}$  and  $\sigma_{\mathfrak{a}}^{-1} \mathrm{Stab}_{\Gamma_0(q)}(\mathfrak{a}) \sigma_{\mathfrak{a}} = \Gamma_\infty$ , but in contrast to the definition given in [Iw3, (2.15)] does *not* in general have determinant 1. Note also that the Definition 6.4 of a scaling matrix is more restrictive than the definition in loc. cit.—Iwaniec’s definition would allow us to multiply  $\tau$  on the right by any  $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ ,  $x \in \mathbb{R}$ .

For  $\mathfrak{a} \in \mathcal{C}(q)$  and  $\sigma_{\mathfrak{a}}$  a scaling matrix, a vector  $\phi \in S_*(q, \chi)$  or  $\mathcal{E}_{it}(q, \chi)$  admits a Fourier expansion at  $\mathfrak{a}$  of the shape

$$(6.8) \quad \phi(\sigma_{\mathfrak{a}} \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}) = (y^{\frac{k}{2}} f)|_{\sigma_{\mathfrak{a}}}(x + iy) = c_{\phi, \mathfrak{a}}(y) + \sum_{n \neq 0} \frac{\rho_{\mathfrak{a}}(n)}{|n|^{1/2}} W(ny) e(nix),$$

where  $W \in \mathcal{W}$  is given by the table in Theorem 6.1(2) and  $c_{\phi, \mathfrak{a}}(y)$  is a possibly-zero constant term. In particular, a classical cusp form  $f$  or non-holomorphic Eisenstein series admits a Fourier expansion of the form (6.8). Sometimes we write  $\rho_{\phi, \mathfrak{a}}(n)$  or  $\rho_{f, \mathfrak{a}}(n)$  for  $\rho_{\mathfrak{a}}(n)$  if we want to emphasize the dependence of  $\rho_{\mathfrak{a}}$  on  $\phi$  or  $f$ .

**Definition 6.5.** *The coefficients  $\rho_{\mathfrak{a}}(n)$  appearing (6.8) are called the Fourier coefficients of  $f$  at the cusp  $\mathfrak{a}$  and depend on the choice of scaling matrix  $\sigma_{\mathfrak{a}}$ .*

The Fourier coefficients  $\rho_{\mathfrak{a}}(n)$  are given in terms of the local Whittaker models at primes dividing  $nq$  (see e.g. [NPS, §3.4.2]). The Fourier coefficients  $\rho_{\mathfrak{a}}(n)$  may also be more explicitly computed in terms of the Hecke eigenvalues  $\lambda_{\pi}(n)$  and other invariants of  $\pi$  using the Jacquet-Langlands local functional equations at primes dividing  $q$ .

*Example.* Consider the cusps  $\infty$  and  $0$ , and choose  $\sigma_{\infty} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$  and  $\sigma_0 = \begin{pmatrix} & -1 \\ q & \end{pmatrix}$ . Then for a newvector  $\phi$  of conductor  $q$  we have

$$(6.9) \quad \rho_0(n) = \epsilon(\pi)_{\text{fin}} \overline{\rho_{\infty}(n)},$$

where  $\epsilon(\pi)_{\text{fin}}$  is the finite root number of the representation  $\pi$ . It satisfies  $|\epsilon(\pi)_{\text{fin}}| = 1$ . The relation (6.9) also follows quickly from [AtLe, Thm. 3(iii)], [Li1, p. 296].

Warning: the coefficients  $\rho_{\mathfrak{a}}(n)$  are in general not multiplicative, nor do they even satisfy the weaker condition  $\rho_{\mathfrak{a}}(nm)\rho_{\mathfrak{a}}(1) = \rho_{\mathfrak{a}}(n)\rho_{\mathfrak{a}}(m)$  for pairs of coprime integers  $m, n$ .

#### 6.4. Kloosterman sums at arbitrary cusps.

**Definition 6.6.** Let  $\mathfrak{a}, \mathfrak{b} \in \mathcal{C}(q)$  and  $\sigma_{\mathfrak{a}}, \sigma_{\mathfrak{b}}$  be scaling matrices. The set

$$\mathcal{C}_{\mathfrak{a}\mathfrak{b}} = \{c > 0 : \begin{pmatrix} * & * \\ c & * \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}\}$$

is called the set of allowed moduli.

Our change in definition of scaling matrices compared to [Iw3] also causes an alteration of the definition of the set of allowed moduli, as well as the Kloosterman sum discussed below. As a consequence, the new definition has the advantage that  $\mathcal{C}_{\mathfrak{a}\mathfrak{b}} \subseteq \mathbb{N}$  for any cusps  $\mathfrak{a}, \mathfrak{b}$ . To help the reader translate between Definition 6.6 and [Iw3], temporarily define  $\mathcal{C}_{\mathfrak{a}\mathfrak{b}}^{\mathbb{Z}}$  to be defined as above, and let  $\mathcal{C}_{\mathfrak{a}\mathfrak{b}}^{\text{Iw}}$  be as in [Iw3]. Then  $\mathcal{C}_{\mathfrak{a}\mathfrak{b}}^{\mathbb{Z}} = (w_{\mathfrak{b}}w_{\mathfrak{a}}^{-1})^{1/2}\mathcal{C}_{\mathfrak{a}\mathfrak{b}}^{\text{Iw}}$ .

*Example.* Take  $\mathfrak{a} = \infty$ ,  $\mathfrak{b} = 0$ ,  $\sigma_{\infty} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ , and  $\sigma_0 = \begin{pmatrix} & -1 \\ q & \end{pmatrix}$ . Then

$$\mathcal{C}_{\infty 0} = \{cq : (c, q) = 1, c \geq 1\}.$$

Let  $\chi$  be an even Dirichlet character modulo  $q$ . For  $\mathfrak{a} \in \mathcal{C}(q)$  and  $\sigma_{\mathfrak{a}}$  a scaling matrix, let  $u_{\mathfrak{a}} \in \Gamma_0(q)$  be such that  $\sigma_{\mathfrak{a}}^{-1}u_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ .

**Definition 6.7.** If a Dirichlet character  $\chi$  modulo  $q$  satisfies  $\chi(u_{\mathfrak{a}}) = 1$ , then we say that  $\mathfrak{a}$  is singular for  $\chi$ .

**Definition 6.8.** If  $\mathfrak{a}, \mathfrak{b}$  are singular cusps for  $\chi$ , then the sum

$$(6.10) \quad S_{\mathfrak{a}\mathfrak{b}}(m, n; c; \chi) = \sum_{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_{\infty} \backslash \sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}/\Gamma_{\infty}} \overline{\chi}(\sigma_{\mathfrak{a}}\gamma\sigma_{\mathfrak{b}}^{-1}) e\left(\frac{am + dn}{c}\right)$$

is called the Kloosterman sum attached to the cusps  $\mathfrak{a}, \mathfrak{b}$ .

If  $|c| \notin \mathcal{C}_{\mathfrak{a}\mathfrak{b}}$ , then the sum appearing in (6.10) is empty, hence  $S_{\mathfrak{a}\mathfrak{b}}(m, n; c; \chi) = 0$ . Temporarily denote by  $S_{\mathfrak{a}\mathfrak{b}}^{\mathbb{Z}}$  the Kloosterman sum appearing in (6.10) and by  $S_{\mathfrak{a}\mathfrak{b}}^{\text{Iw}}$  the sum appearing in e.g. [Iw3, (3.13)]. If  $\chi$  is an even Dirichlet character modulo  $q$ , then  $S_{\mathfrak{a}\mathfrak{b}}^{\mathbb{Z}}$  and  $S_{\mathfrak{a}\mathfrak{b}}^{\text{Iw}}$  are related by

$$(6.11) \quad S_{\mathfrak{a}\mathfrak{b}}^{\mathbb{Z}}(m, n, c; \chi) = S_{\mathfrak{a}\mathfrak{b}}^{\text{Iw}}(m, n, \sqrt{w_{\mathfrak{a}}w_{\mathfrak{b}}^{-1}}c; \chi).$$

*Example.* Take  $\mathfrak{a} = \infty$ ,  $\mathfrak{b} = 0$ ,  $\sigma_{\infty} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ , and  $\sigma_0 = \begin{pmatrix} & -1 \\ q & \end{pmatrix}$ . Then (see [KY2, (2.20)])

$$(6.12) \quad S_{\infty 0}(m, n; cq; \chi) = \overline{\chi}(c)S(\overline{q}m, n; c).$$

**6.5. The Bruggeman-Kuznetsov formula.** Let

$$V(q) = \text{Vol}(\Gamma_0(q)\backslash\mathcal{H}) = \frac{\pi}{3}q \prod_{p|q} (1+p^{-1}),$$

choose  $\Phi \in C_c^\infty(\mathbb{R}_{>0})$ ,  $\chi$  an even Dirichlet character modulo  $q$ , singular cusps  $\mathbf{a}, \mathbf{b}$ , and set

$$(6.13) \quad \mathcal{K} = \sum_{c \in \mathcal{C}_{\mathbf{a}\mathbf{b}}} S_{\mathbf{a}\mathbf{b}}(m, n; c; \chi) \Phi\left(\sqrt{\frac{w_{\mathbf{a}}}{w_{\mathbf{b}}}}c\right).$$

Define the integral transforms

$$(6.14) \quad 2\pi i^{-k} \mathcal{L}^{\text{hol}}\Phi(k) = \frac{1}{2\pi i} \int_{(1)} \frac{2^s \Gamma\left(\frac{s+k-1}{2}\right)}{\Gamma\left(\frac{k+1-s}{2}\right)} \tilde{\Phi}(s+1) (4\pi\sqrt{|mn|})^{-s} ds,$$

and

$$(6.15) \quad \mathcal{L}^\pm \Phi(t) = \frac{1}{2\pi i} \int_{(2)} h_\pm(s, t) \tilde{\Phi}(s+1) (4\pi\sqrt{|mn|})^{-s} ds,$$

where

$$(6.16) \quad h_\pm(s, t) = \frac{2^s}{2\pi^2} \Gamma\left(\frac{s}{2} + it\right) \Gamma\left(\frac{s}{2} - it\right) \begin{cases} \cos(\pi s/2), & \pm = + \\ \cosh(\pi t), & \pm = -, \end{cases}$$

and  $\tilde{\Phi}(s) = \int_0^\infty \Phi(x) x^s \frac{dx}{x}$  denotes the Mellin transform of  $\Phi$ .

**Theorem 6.9** (Bruggeman-Kuznetsov Formula). *Let  $\Phi \in C_c^\infty(\mathbb{R}_{>0})$  and  $\mathcal{K}$  be as in (6.13). We have*

$$\mathcal{K} = \mathcal{K}_{\text{Maass}} + \mathcal{K}_{\text{Eis}} + \mathcal{K}_{\text{hol}},$$

where

$$(6.17) \quad \mathcal{K}_{\text{Maass}} = \frac{4\pi}{V(q)} \sum_{t_j} \mathcal{L}^\pm \Phi(t_j) \sum_{\epsilon=0,1} (\pm 1)^\epsilon \sum_{\substack{f \in \mathcal{B}_{it_j}(q, \chi) \\ \text{of parity } \epsilon}} \overline{\rho_{f, \mathbf{a}}(m)} \rho_{f, \mathbf{b}}(n),$$

$$(6.18) \quad \mathcal{K}_{\text{Eis}} = \frac{4\pi}{V(q)} \int_{-\infty}^{\infty} \mathcal{L}^\pm \Phi(t) \sum_{\epsilon=0,1} (\pm 1)^\epsilon \sum_{\substack{E \in \mathcal{B}_{it, \text{Eis}}(q, \chi) \\ \text{of parity } \epsilon}} \overline{\rho_{E, \mathbf{a}}(m)} \rho_{E, \mathbf{b}}(n) dt,$$

where one takes  $+$  in  $\pm$  (resp.  $-$ ) if  $mn > 0$  (resp.  $mn < 0$ ), and

$$(6.19) \quad \mathcal{K}_{\text{hol}} = \frac{4\pi}{V(q)} \sum_{k>0, \text{ even}} \mathcal{L}^{\text{hol}}\Phi(k) \sum_{f \in \mathcal{B}_k(q, \chi)} \overline{\rho_{f, \mathbf{a}}(m)} \rho_{f, \mathbf{b}}(n)$$

if  $mn > 0$ , and  $\mathcal{K}_{\text{hol}} = 0$  if  $mn < 0$ .

Above,  $\mathcal{B}_*(q, \chi)$  denotes any orthonormal basis of  $S_*(q, \chi)$  with respect to the probability measure on  $\Gamma_0(q)\backslash\mathcal{H}$ , and  $\mathcal{B}_{it, \text{Eis}}(q, \chi)$  denotes an orthonormal basis of  $\mathcal{E}_{it}(q, \chi)$  with respect to the formal inner product divided by  $V(q)$  [PY2, §2.2].

Remark: The above formula is taken from [KY1, Thm. 3.5] and [Y2, (10.2)] (which contains a typo: the factor of  $4\pi$  on the right hand side should be deleted), but has been normalized differently in two ways. First, we have defined the Fourier coefficients  $\rho_{\mathbf{a}}(n)$  using the canonical normalization of archimedean Whittaker models chosen in Theorem

6.1. One of the consequences of this choice is that  $\rho_{\mathfrak{a}}(n) = \rho_{\mathfrak{a}}(-n)$  by definition (the factor  $(\operatorname{sgn} n)^\epsilon$  is naturally part of the archimedean Whittaker function). Explicitly, we have  $\nu_{\mathfrak{a}}(n) = 2(\operatorname{sgn} n)^\epsilon \rho_{\mathfrak{a}}(n)$ , where  $\nu_{\mathfrak{a}}(n)$  is defined by [Iw4, (8.5)]. Secondly, we have chosen probability measure on  $\Gamma_0(q)\backslash\mathcal{H}$  to define inner products, whereas most authors choose the push-forward measure from  $\mathcal{H}$ . These two choices result in the appearance of the factor of  $\frac{V(q)}{4\pi}$ , which is natural, it being the leading constant in Weyl's law for  $\Gamma_0(q)\backslash\mathcal{H}$ .

6.6. **Explicit choice of basis.** Let  $\mathcal{H}_{it_j}(m, \chi)$  be the (finite) set of cuspidal automorphic representations  $\pi$  with conductor  $\mathfrak{c}(\pi) = m$ , finite order central character  $\omega_\pi = \chi$ , and  $\pi_\infty \simeq \pi(\alpha^{it_j} \operatorname{sgn}^\epsilon, \alpha^{-it_j} \operatorname{sgn}^\epsilon)$ . One may alternatively (and equivalently) take  $\mathcal{H}_{it_j}(m, \chi)$  as in [PY2, §2.1] to be the set of cuspidal Hecke-Maass newforms of level  $m$ , spectral parameter  $t_j$ , and central character  $\chi$ .

Similarly, let  $\mathcal{H}_k(m, \chi)$  be the (finite) set of cuspidal automorphic representations  $\pi$  with  $\mathfrak{c}(\pi) = m$ , finite order  $\omega_\pi = \chi$  and  $\pi_\infty \simeq \sigma(\chi_1, \chi_2)$  with  $\chi_1 \chi_2^{-1} = \alpha^s \operatorname{sgn}^m$  for some  $s - m \in 1 + 2\mathbb{Z}_{\geq 0}$  satisfying  $s + 1 = k$ . One may also just as well take  $\mathcal{H}_k(m, \chi)$  as in [PY2, §2.1] to be the set of cuspidal holomorphic newforms of level  $m$ , weight  $k$ , and central character  $\chi$ . All statements that follow involving  $\mathcal{H}_*(m, \chi)$  will hold equally well with either definition.

Finally, let  $\mathcal{H}_{it, \text{Eis}}(m, \chi)$  be the (finite) set of pairs  $(\mu_1, \mu_2)$  of unitary Hecke characters of  $\mathbb{Q}$  such that the global principal series representation  $\pi = \pi(\mu_1, \mu_2)$  (see e.g. [B, §3.7]) has  $\mathfrak{c}(\pi) = m$ ,  $\mu_1 \mu_2 = \chi$  of finite order, and  $\pi_\infty \simeq \pi(\alpha^{it} \operatorname{sgn}^\epsilon, \alpha^{-it} \operatorname{sgn}^\epsilon)$ . One may also take  $\mathcal{H}_{it, \text{Eis}}(m, \chi)$  to be the set of newform Eisenstein series of level  $m$  and character  $\chi$  as defined in [PY2, §2.2]. Using the notation of loc. cit. §2.2, the bijection between these two definitions for  $\mathcal{H}_{it, \text{Eis}}(m, \chi)$  is given by

$$(\mu_1, \mu_2) \mapsto E_{\chi_1, \chi_2}(z, 1/2 + it),$$

where  $\chi_1$  and  $\chi_2$  are the primitive Dirichlet characters corresponding to  $\chi_1 = \mu_1|_{\widehat{\mathbb{Z}}^\times}$  and  $\chi_2 = \overline{\mu_1}|_{\widehat{\mathbb{Z}}^\times}$ . Either definition will make sense in what follows.

If  $\chi = 1$  is trivial, we may use the shorthand  $\mathcal{H}_*(m) := \mathcal{H}_*(m, 1)$ , as well as the shorthand  $\mathcal{H}_* := \bigcup_m \mathcal{H}_*(m)$ , where  $*$  =  $it_j$ ,  $k$ , or  $it, \text{Eis}$ .

For  $(\pi, V)$  a cuspidal representation of conductor  $m$  and central character  $\omega_\pi$ , write

$$\pi^{K_0(m\ell)} = \{\phi \in V : \pi(g)\phi = \omega_\pi(g)\phi \text{ for all } g \in K_0(m\ell)\}.$$

The set of fixed vectors  $\pi^{K_0(m\ell)}$  is also called an oldclass in the classical terminology, i.e.

$$S_*(\ell, f, \chi) = \pi^{K_0(m\ell)}, \quad \text{via } f \mapsto \phi_f,$$

where  $S_*(\ell, f, \chi)$  was the notation used in [PY2, (2.5)].

As a first step in the construction of an orthonormal basis for  $S_*(q, \chi)$ , observe that forms  $\phi \in S_*(q, \chi)$  that generate distinct irreducible cuspidal automorphic representations are necessarily orthogonal to each other. Thus, we have the orthogonal direct sum

$$(6.20) \quad S_*(q, \chi) = \bigoplus_{m\ell=q} \bigoplus_{\pi \in \mathcal{H}_*(m, \chi)} \pi^{K_0(m\ell)}.$$

By (6.20), the problem of choosing an orthonormal basis for  $S_*(q, \chi)$  reduces to choosing orthonormal bases for the oldclasses  $\pi^{K_0(m\ell)}$ .

Write  $\phi_d$  for the function  $g \mapsto \phi((\begin{smallmatrix} d & \\ & 1 \end{smallmatrix})g)$  with  $(\begin{smallmatrix} d & \\ & 1 \end{smallmatrix}) \in \operatorname{GL}_2(\mathbb{R}) \hookrightarrow \operatorname{GL}_2(\mathbb{A})$  in the first position. If  $\phi$  is a newvector for  $\pi$  of conductor  $m$ , we have by Atkin-Lehner-Li theory that

$$(6.21) \quad \pi^{K_0(m\ell)} = \operatorname{span}\{\phi_d : d \mid \ell\}.$$

We write an orthonormal basis  $\mathcal{B}(\ell, \pi)$  for  $\pi^{K_0(m\ell)}$  in the coordinates (6.21) as

$$\mathcal{B}(\ell, \pi) = \{\phi^{(\delta)} = \sum_{d|\ell} x_\delta(d) \phi_d : \delta \mid \ell\}$$

for some choice of coefficients  $x_\delta(d)$ . Thus an orthonormal basis for  $S_*(q, \chi)$  is given by

$$\mathcal{B}_*(q, \chi) = \bigcup_{m\ell=q} \bigcup_{\pi \in \mathcal{H}_*(m, \chi)} \{\phi^{(\delta)} : \phi \text{ newvector for } \pi, \delta \mid \ell\}.$$

Taking the Fourier expansion of the newvector  $\phi$  at  $\infty$  as in Theorem 6.1, we have that the Fourier coefficients at infinity of the forms  $\phi_d$  and  $\phi^{(\delta)}$  are related by

$$(6.22) \quad \rho_{\phi_d}(n) = d^{1/2} \rho_\phi(n/d), \quad \text{and} \quad \rho_{\phi^{(\delta)}}(n) = \sum_{d|\ell} d^{1/2} x_\delta(d) \rho_\phi(n/d),$$

where if  $n/d$  is not an integer, we interpret  $\rho_\phi(n/d) = 0$ . Since  $\rho_\phi(n)$  are directly related to Hecke eigenvalues via (6.4), we also define

$$(6.23) \quad \lambda_\pi^{(\delta)}(n) = \sum_{d|\ell} d^{1/2} x_\delta(d) \lambda_\pi(n/d),$$

where likewise  $\lambda_\pi(n/d) = 0$  if  $n/d$  is not an integer. Note that we have  $\rho_{\phi^{(\delta)}}(n) = \rho_\phi(1) \lambda_\pi^{(\delta)}(n)$ .

We denote by  $\epsilon(\pi)_{-1}$  the parity of  $\pi$  (in line with Iwaniec's notation  $T_{-1}$  for the involution  $f(z) \mapsto f(-\bar{z})$  on Hecke-Maass forms). Lastly, we set  $\epsilon_\pi^{(\pm)} = (\pm 1)^{\epsilon(\pi)-1} \epsilon(\pi)_{\text{fin}}$ , where  $\epsilon(\pi)_{\text{fin}}$  was defined in (6.9). With these notations we have the following.

**Theorem 6.10** (Explicit Bruggeman-Kuznetsov Formula for cusps  $\infty, 0$ ). *Let  $\Phi \in C_c^\infty(\mathbb{R}_{>0})$ . We have*

$$(6.24) \quad \mathcal{K} = \sum_{(c,q)=1} \bar{\chi}(c) S(\bar{q}m, n; c) \Phi(q^{1/2}c) = \mathcal{K}_{\text{Maass}} + \mathcal{K}_{\text{Eis}} + \mathcal{K}_{\text{hol}},$$

with notation as follows. We have

$$(6.25) \quad \mathcal{K}_{\text{Maass}} = \frac{4\pi}{V(q)} \sum_{t_j} \mathcal{L}^\pm \Phi(t_j) \sum_{\ell r=q} \sum_{\pi \in \mathcal{H}_{it_j}(r, \chi)} \frac{\epsilon_\pi^{(\pm)}}{\mathcal{L}_\pi^*(1)} \sum_{\delta|\ell} \bar{\lambda}_\pi^{(\delta)}(|m|) \bar{\lambda}_\pi^{(\delta)}(|n|),$$

$$(6.26) \quad \mathcal{K}_{\text{Eis}} = \frac{4\pi}{V(q)} \int_{-\infty}^{\infty} \mathcal{L}^\pm \Phi(t) \sum_{\ell r=q} \sum_{\pi \in \mathcal{H}_{it, \text{Eis}}(r, \chi)} \frac{\epsilon_\pi^{(\pm)}}{2\pi \mathcal{L}_\pi^*(1)} \sum_{\delta|\ell} \bar{\lambda}_\pi^{(\delta)}(|m|) \bar{\lambda}_\pi^{(\delta)}(|n|) dt,$$

where one takes  $+$  in  $\pm$  (resp.  $-$ ) if  $mn > 0$  (resp.  $mn < 0$ ), and

$$(6.27) \quad \mathcal{K}_{\text{hol}} = \frac{4\pi}{V(q)} \sum_{k>0, \text{ even}} \mathcal{L}^{\text{hol}} \Phi(k) \sum_{\ell r=q} \sum_{\pi \in \mathcal{H}_k(r, \chi)} \frac{\epsilon(\pi)_{\text{fin}}}{\mathcal{L}_\pi^*(1)} \sum_{\delta|\ell} \bar{\lambda}_\pi^{(\delta)}(|m|) \bar{\lambda}_\pi^{(\delta)}(|n|)$$

if  $mn > 0$ , and  $\mathcal{K}_{\text{hol}} = 0$  if  $mn < 0$ .

There are various choices of basis for  $\pi^{K_0(m\ell)}$  in the literature (see e.g. [ILS] [PY1] [BM]), and it is not clear that there is any canonical choice for general level. Let  $\xi_\delta(d)$  be the coefficients defined in [P2, Prop. 7.1]. The choice  $x_\delta(d) = \xi_\delta(d)$  defines an orthonormal basis  $\{\phi^{(\delta)} : \delta \mid \ell\}$  for  $\pi^{K_0(m\ell)}$  (see [SPY, Thm. 3.2] for a nice proof that avoids the Rankin-Selberg method). The coefficients  $\xi_\delta(d)$  are given in terms of the divisors of  $d$  and  $\delta$  and the Hecke eigenvalues of  $\pi$ . Inspecting the definition of  $\xi_\delta(d)$ , one deduces the following lemma.

**Lemma 6.11.** *The coefficients  $\xi_\delta(d)$  enjoy the following properties:*

- (1) *The coefficients  $\xi_\delta(d)$  are supported on  $d|\delta$ .*
- (2) *The function  $\xi_\delta(d)$  is jointly multiplicative in  $\delta, d$ .*
- (3) *We have  $\xi_\delta(d) \ll (\delta d)^\varepsilon$ .*

As a consequence of Lemma 6.11(1), the  $\lambda_\pi^{(\delta)}(n)$  associated to  $\xi_\delta(d)$  (see (6.23)) is jointly multiplicative in  $\delta, n$  since it is the Dirichlet convolution of jointly multiplicative functions. The coefficients  $\xi_\delta(d)$  also give an orthonormal basis in the case of the Eisenstein series (see [Y2, §8] for details).

## 7. TOOLS FROM ANALYTIC NUMBER THEORY

**7.1. Gauss sums.** We will need estimates for Gauss sums of non-primitive Dirichlet characters.

**Lemma 7.1.** *Let  $\chi$  be a Dirichlet character modulo  $q$ , induced by the primitive character  $\chi'$  modulo  $q'$ . For  $n \in \mathbb{Z}$ , let*

$$\tau(\chi, n) = \sum_{x \pmod{q}} \chi(x) e_q(nx).$$

Then

$$(7.1) \quad \tau(\chi, n) = \tau(\chi') \sum_{d|(n, q/q')} d \overline{\chi'}\left(\frac{n}{d}\right) \chi'\left(\frac{q}{dq'}\right) \mu\left(\frac{q}{dq'}\right).$$

In particular,  $\tau(\chi) = \tau(\chi, 1) = \mu(q/q') \chi'(q/q') \tau(\chi')$ . Moreover, if  $\chi$  is any Dirichlet character modulo  $q$ , induced by  $\chi'$  modulo  $q'$  (including the trivial character with  $q' = 1$ ), we have

$$(7.2) \quad |\tau(\chi, n)| \leq (q')^{1/2} \left(n, \frac{q}{q'}\right).$$

Remark. [IK, Lem. 3.2] is relevant but has misprints, so we have included a proof.

*Proof.* By Möbius inversion,

$$\tau(\chi, n) = \sum_{d|\frac{q}{q'}} \mu\left(\frac{q}{dq'}\right) \chi'\left(\frac{q}{dq'}\right) \sum_{y \pmod{q'd}} \chi'(y) e_{q'd}(ny).$$

Changing variables  $y \rightarrow y + q'$  shows that the inner sum over  $y$  vanishes unless  $d|n$ , in which case the sum over  $y$  is a Gauss sum for  $\chi'$  repeated  $d$  times. It is well-known that

$$\sum_{y \pmod{q'}} \chi'(y) e_{q'}(my) = \overline{\chi'}(m) \tau(\chi'),$$

valid for all  $m \in \mathbb{Z}$ . This gives (7.1). Finally, (7.2) follows easily from (7.1).  $\square$

**Corollary 7.2.** *Suppose  $\chi$  is a character of prime power modulus  $q = p^\beta$ ,  $\beta \geq 1$  and conductor  $q'$ . Let  $n$  be an integer. Then  $\tau(\chi) \tau(\chi, n) = 0$  except when the following conditions hold:*

- (1) *If  $q' = q$  and  $(n, q) = 1$ .*
- (2) *If  $q' = 1$  and  $q = p$ .*

*Proof.* If  $1 < q' < q$ , then  $\tau(\chi) = 0$ . If  $q' = q$ , then  $\tau(\chi, n) = 0$  unless  $(n, q) = 1$ . If  $q' = 1$  then  $\tau(\chi) = S(1, 0; q)$ , which vanishes unless  $q = p$ .  $\square$

## 7.2. Approximate functional equation for a divisor function times a character.

There are various ways to solve a shifted convolution/divisor problem, including the circle method, the delta symbol method, and via inner products with Poincare series. Here we prove a generalized form of [Y1, Lem. 5.4], which will be convenient for our purposes.

**Lemma 7.3.** *Let  $\chi$  be a primitive Dirichlet character modulo  $q$ . Let  $G(s)$  be an even entire holomorphic function with rapid decay in vertical strips, satisfying  $G(0) = 1$  (e.g.  $G(s) = \exp(s^2)$ ). Then*

$$(7.3) \quad \tau(n)\chi(n) = \frac{2}{\tau(\bar{\chi})} \sum_{c=1}^{\infty} \frac{\chi(c)}{c} f\left(\frac{c}{\sqrt{n}}\right) \sum_{r \pmod{cq}}^* \bar{\chi}(r) e_{cq}(nr),$$

where

$$f(x) = \frac{1}{2\pi i} \int_{(1)} x^{-2s} L(1+2s, \chi_{0,q}) \frac{G(s)}{s} ds,$$

and where  $\chi_{0,q}$  denotes the trivial character modulo  $q$ .

Remarks.

- (1) The proof of Lemma 7.3 gives an even more general formula than (7.3).
- (2) It turns out to be highly convenient that  $c$  runs over integers coprime to  $q$ .
- (3) It is not hard to check that  $f(x)$  is smooth for  $x > 0$ , and satisfies the bound

$$(7.4) \quad x^j f^{(j)}(x) \ll_{j,\varepsilon,A} \frac{x^{-\varepsilon} q^\varepsilon}{(1+x)^A}.$$

*Proof.* For Dirichlet characters  $\chi_1, \chi_2$  to moduli  $q_1, q_2$  respectively, define

$$\lambda_{\chi_1, \chi_2}(n, s) := \sum_{ab=n} \chi_1(a) \bar{\chi}_2(b) \left(\frac{b}{a}\right)^{s-\frac{1}{2}},$$

where the notation matches that in [Y2]. Observe  $\chi(n)\tau(n) = \lambda_{\chi, \bar{\chi}}(n, 1/2)$ , and note the functional equation

$$(7.5) \quad \lambda_{\chi_1, \chi_2}(n, 1-s) = \lambda_{\bar{\chi}_2, \bar{\chi}_1}(n, s).$$

Now suppose  $\chi_1, \chi_2$  are primitive, and observe that

$$\lambda_{\chi_1, \chi_2}(n, s) = \frac{n^{s-\frac{1}{2}}}{\tau(\chi_2)} \sum_{c=1}^{\infty} \frac{\chi_1(c)}{c^{2s}} \sum_{r \pmod{cq_2}} \chi_2(r) e_{cq_2}(nr),$$

by splitting the sum over  $r$  into residue classes modulo  $q_2$ . Next we factor out  $a = \gcd(c, r)$  and change variables  $c \rightarrow ac$  and  $r \rightarrow ar$ , giving

$$(7.6) \quad \lambda_{\chi_1, \chi_2}(n, s) = \frac{n^{s-\frac{1}{2}}}{\tau(\chi_2)} L(2s, \chi_1 \chi_2) \sum_{c=1}^{\infty} \frac{\chi_1(c)}{c^{2s}} \sum_{r \pmod{cq_2}}^* \chi_2(r) e_{cq_2}(nr).$$

Consider

$$\frac{1}{2\pi i} \int_{(1)} \lambda_{\chi_1, \chi_2}(n, s + \frac{1}{2}) \frac{G(s)}{s} ds.$$

Shifting the contour to  $\operatorname{Re}(s) = -1$ , applying (7.5), and changing variables  $s \rightarrow -s$  gives

$$(7.7) \quad \lambda_{\chi_1, \chi_2}(n, 1/2) = \frac{1}{2\pi i} \int_{(1)} \lambda_{\chi_1, \chi_2}(n, s + \frac{1}{2}) \frac{G(s)}{s} ds + \frac{1}{2\pi i} \int_{(1)} \lambda_{\bar{\chi}_2, \bar{\chi}_1}(n, s + \frac{1}{2}) \frac{G(s)}{s} ds.$$



Since both  $\chi_1, \chi_2$  are primitive, we may insert (7.6) into the two integrals. The first term in (7.7) then equals

$$\frac{1}{\tau(\chi_2)} \sum_{c=1}^{\infty} \frac{\chi_1(c)}{c} \sum_{r \pmod{cq_2}}^* \chi_2(r) e_{cq_2}(nr) f\left(\frac{c}{\sqrt{n}}\right).$$

where

$$f(x) = \frac{1}{2\pi i} \int_{(1)} x^{-2s} L(1+2s, \chi_1\chi_2) \frac{G(s)}{s} ds,$$

and the second term is similar. The lemma then follows, taking  $\chi_1 = \chi$ ,  $\chi_2 = \bar{\chi}$ .  $\square$

**7.3. The large sieve inequality.** Let us denote by

$$(7.8) \quad \int_{*\leq T} \text{any of } \sum_{|t_j|\leq T}, \sum_{k\leq T}, \text{ or } \int_{|t|\leq T} dt$$

according to whether  $* = it_j, k$ , or  $it$ , Eis.

**Lemma 7.4** (Spectral large sieve). *For any sequence of complex numbers  $a_n$ , we have*

$$\int_{*\leq T} \sum_{\pi \in \mathcal{H}_*(q)} \left| \sum_{n \leq N} a_n \lambda_{\pi}(n) \right|^2 \ll_{\varepsilon} (T^2 q + N) (qTN)^{\varepsilon} \sum_{n \leq N} |a_n|^2.$$

*Proof.* The spectral sum on the left hand side is estimated by [DI, Thm. 2], but with weights  $\mathcal{L}_{\pi}^*(1)^{-1}$ . These weights may be absorbed into the factor  $(qT)^{\varepsilon}$  on the right hand side.  $\square$

**7.4. Additional spectral bounds.**

**Lemma 7.5.** *Suppose  $(q_1, q_2) = 1$  and  $(n, q_1 q_2) = 1$ . Then*

$$(7.9) \quad \sum_{|t_j|\leq T} \sum_{\psi \pmod{q_2}}^+ \sum_{\pi \in \mathcal{H}_{it_j}(q_1 q_2, \psi)} |\lambda_{\pi}(n)|^2 \ll (T^2 q_1 q_2^2 + n^{1/2} q_2^{1/2}) (nT q_1 q_2)^{\varepsilon},$$

where the  $+$  indicates that the sum runs over even Dirichlet characters  $\psi$ .

*Remarks.* The case  $q_1 = q_2 = 1$  can be found in [Mo, Lem. 2.4], and the case  $q_2 = 1$  is a special case of [BM, Lem. 12]. The idea is to use the Bruggeman-Kuznetsov formula together with the Weil bound. We have not stated the analogous bounds for holomorphic forms or Eisenstein series, since these cases follow immediately from  $|\lambda_f(n)| \leq \tau(n)$  which is Deligne's bound in the holomorphic case, and directly established for Eisenstein series.

*Proof.* Weighting by  $\mathcal{L}_{\pi}^*(1)$  and extending the newforms to an orthogonal basis of  $S_{it_j}(q_1 q_2, \psi)$  in an arbitrary way, we have by (6.3), (6.4) and positivity that the left hand side of (7.9) is

$$(7.10) \quad \ll (q_1 q_2 T)^{o(1)} \sum_{|t_j|\leq T} \sum_{\psi \pmod{q_2}}^+ \sum_{\phi \in \mathcal{B}_{it_j}(q_1 q_2, \psi)} |\rho_{\phi}(n)|^2.$$

Next we extend the sum over  $t_j$  in (7.10) to the whole spectrum and insert the following smooth weights. To capture  $t_j \ll (T q_1 q_2)^{\varepsilon}$ , we attach the weight function  $h_V(t) = (t^2 + 1/4) \exp(-t^2/V^2)$ , with  $V = (T q_1 q_2)^{\varepsilon}$  to the spectrum. For  $t_j \gg (T q_1 q_2)^{\varepsilon}$  we attach a sum of weights of the form  $h_{U,V}(t) = \sum_{\pm} \exp(-(\pm t - U)^2/V^2)$ , with  $(T q_1 q_2)^{\varepsilon} \ll U \ll T$

and  $V = U^{1-\varepsilon}$ . We then apply the Bruggeman-Kuznetsov formula, showing that (7.10) is bounded by a sum of expressions of the form

$$\sum_{\psi \pmod{q_2}}^+ (UVq_1q_2 + q_1q_2K_\psi)(q_1q_2T)^\varepsilon,$$

where

$$K_\psi = \sum_{c \equiv 0 \pmod{q_1q_2}} c^{-1} S_\psi(n, n; c) B\left(\frac{4\pi n}{c}\right), \quad S_\psi(m, n; c) = \sum_{y \pmod{c}} \bar{\psi}(y) e_c(ym + \bar{y}n),$$

and  $B(x)$  is the Bessel transform of either  $h_V$  or  $h_{U,V}$  that appears in the Bruggeman-Kuznetsov formula (see [Iw4, (9.10)]).

For a spectral weight function of the type  $h_{U,V}$ , Jutila and Motohashi [JM2] showed the bound  $B(x) \ll x^{-1/2}U^2$  for  $x \gg U^{2+\varepsilon}$ , and that  $B(x)$  is very small otherwise. For the case of  $h_V$ , one may also easily show two crude bounds as follows. One simple bound is  $B(x) \ll V^4$ , using the easy bound  $\frac{|J_{2it}(x)|}{\cosh(\pi t)} \ll 1$  and which follows from the integral representation [GR, 8.411.4]. Hence  $B(x) \ll V^4 \ll V(q_1q_2T)^\varepsilon$ . We also claim  $B(x) \ll xV^C$  for some fixed  $C > 0$ , which can be derived by shifting contours to the line  $\operatorname{Re}(2it) = 1$ , in the integral representation [Iw4, (9.10)], and bounding the integral trivially (one can find more details in [PY2, Pf. of Lem. 10.2]). Altogether, we derive the bound

$$(7.11) \quad B(x) \ll (q_1q_2T)^\varepsilon \min\left(V, \frac{x}{V}\right),$$

valid for both classes of test functions  $h_{U,V}$  or  $h_V$ .

It suffices to bound the contribution from  $K_\psi$ . We have

$$K := \sum'_{\psi \pmod{q_2}} K_\psi = \sum_{c \equiv 0 \pmod{q_1q_2}} c^{-1} B\left(\frac{4\pi n}{c}\right) \sum_{\psi \pmod{q_2}} \frac{1}{2}(1 + \psi(-1)) \sum_{y \pmod{c}} \bar{\psi}(y) e_c(yn + \bar{y}n).$$

The sum over  $\psi$  detects the condition  $y \equiv \pm 1 \pmod{q_2}$ , giving

$$K = \frac{1}{2}\varphi(q_2) \sum_{\pm} \sum_{c \equiv 0 \pmod{q_1q_2}} c^{-1} B\left(\frac{4\pi n}{c}\right) \sum_{\substack{y \pmod{c} \\ y \equiv \pm 1 \pmod{q_2}}} e_c(yn + \bar{y}n).$$

Write  $c = c_1c_2$  where  $c_2|q_2^\infty$  and  $(c_1, q_2) = 1$ . We claim

$$(7.12) \quad \sum_{\substack{y \pmod{c} \\ y \equiv \pm 1 \pmod{q_2}}} e_c(yn + \bar{y}n) \ll \tau(c_1)c_1^{1/2}(n, c_1)^{1/2} \frac{c_2}{q_2},$$

as we now show. The sum (7.12) factors as  $S_1S_2$  where

$$S_1 = \sum_{y \pmod{c_1}}^* e_{c_1}(yn\bar{c}_2 + \bar{y}n\bar{c}_2), \quad S_2 = \sum_{\substack{y \pmod{c_2} \\ y \equiv \pm 1 \pmod{q_2}}} e_{c_2}(yn\bar{c}_1 + \bar{y}n\bar{c}_1).$$

By a trivial bound, we have  $S_2 \ll \frac{c_2}{q_2}$ , while  $S_1 = S(n\bar{c}_2, n\bar{c}_2; c_1)$  is the usual Kloosterman sum. The Weil bound completes the proof of the claim.

Therefore, we have

$$K \ll \sum_{\substack{c_2 \equiv 0 \pmod{q_2} \\ c_2 | q_2^\infty}} \sum_{\substack{c_1 \equiv 0 \pmod{q_1} \\ (c_1, q_2) = 1}} c_1^{-1/2+\varepsilon} (n, c_1)^{1/2} \left| B\left(\frac{4\pi n}{c_1 c_2}\right) \right| \ll \frac{n^{1/2+\varepsilon} q_2^{1/2} (q_1 q_2 T)^\varepsilon}{q_1 q_2},$$

using (7.11), which completes the proof.  $\square$

**Theorem 7.6.** *Suppose  $q = q_1 q_2$  with  $(q_1, q_2) = 1$  and  $T \geq 1$ . Then*

$$(7.13) \quad \sum_{\eta \pmod{q_2}} \sum_{|t_j| \leq T} \sum_{\ell m = q} (\ell, q_1) \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} |L(1/2 + it, \pi \otimes \bar{\eta})|^4 \ll_t q_1 q_2^2 T^2 (q_1 q_2 T)^\varepsilon,$$

with polynomial dependence on  $t$ . A similar bound holds true for holomorphic forms, as well as the Eisenstein series.

Remarks. If  $q_2 = 1$ , then this is a ‘‘standard’’ fourth moment bound for automorphic  $L$ -functions, which follows from the spectral large sieve inequality (Lemma 7.4). It is thus the  $q_2$ -aspect that has novelty. Our proof of Theorem 7.6 eventually reduces the problem to the case  $q_2 = 1$ .

To gauge the content of Theorem 7.6, it is helpful to discuss two special cases. First, suppose that  $q_1 = 1$  and  $q_2 = p$ , prime. The main contribution to (7.13) comes from  $m = p$  and  $\eta$  non-trivial, in which case  $\pi \otimes \bar{\eta} \in \mathcal{H}_{it_j}(p^2, 1)$ . On such forms the map  $(\pi, \eta) \rightarrow \pi \otimes \bar{\eta}$  is at most two-to-one (see Lemma 7.7 below), the multiplicity arising from a quadratic twist. Hence Theorem 7.6 follows from the standard fourth moment bound of level  $p^2$ .

Next consider the case  $q_1 = 1$ ,  $q_2 = p^2$ , with  $p$  prime. If  $m = p^2$  and  $\eta$  is primitive modulo  $p^2$ , then  $\pi \otimes \bar{\eta} \in \mathcal{H}_{it_j}(p^4, 1)$ . Again, the multiplicity of the map  $(\pi, \eta) \rightarrow \pi \otimes \bar{\eta}$  is bounded, and the standard level  $p^4$  fourth moment bound suffices to estimate the contribution of these forms to the left hand side of (7.13). Now consider the contribution from  $\eta$  of conductor  $p$ . Consider the typical case that  $\pi$  is twist-minimal with  $m = p^2$ . Then  $\pi \otimes \bar{\eta} \in \mathcal{H}_{it_j}(p^2, 1)$ , which is of lower-level than the previous case. On the other hand, the map  $(\pi, \eta) \rightarrow \pi \otimes \bar{\eta}$  has multiplicity  $\gg p$ , seen as follows. Suppose  $\pi \in \mathcal{H}_{it_j}(p^2, \eta^2)$ , and suppose  $\chi$  has conductor  $p$ . Then  $\pi_\chi := \pi \otimes \chi \in \mathcal{H}_{it_j}(p^2, (\eta\chi)^2)$  and  $\pi \otimes \bar{\eta} = \pi_\chi \otimes \bar{\eta}\bar{\chi}$ , so that there are  $p - 1$  distinct pairs  $(\pi_\chi, \chi\eta)$  all having the same twisted form  $\pi \otimes \bar{\eta}$ . Luckily, the extra multiplicity is compensated by the saving in the number of forms of level  $p^2$  compared to those of level  $p^4$ .

*Proof.* For simplicity of exposition, we only give a proof in the case that  $q_1 = 1$  and  $t = 0$ . The generalization to Theorem 7.6 consists of only notational difficulties.

Abusing notation, for the duration of this proof we denote by  $(\widehat{\mathbb{Z}/q\mathbb{Z}})^\times$  the group of finite-order Hecke characters of  $\mathbb{Q}$  with conductor dividing  $q$ . (For intuition, note also that  $(\widehat{\mathbb{Z}/q\mathbb{Z}})^\times$  is naturally isomorphic to the group of Dirichlet characters modulo  $q$ .) Define

$$(7.14) \quad \mathcal{H}_{* \leq T}^{\text{tw}}(q) := \bigcup_{\eta \in (\widehat{\mathbb{Z}/q\mathbb{Z}})^\times} \bigcup_{* \leq T} \bigcup_{m|q} \{(\pi, \eta) : \pi \in \mathcal{H}_*(m, \eta^2)\},$$

where  $*$  is any of  $it_j$ ,  $k$ , or  $it$ , Eis as in Section 6.6, and  $* \leq T$  denotes either  $|t_j| \leq T$ ,  $k \leq T$ , or  $|t| \leq T$  in each of the three cases of  $*$ , respectively.

If there exists  $\chi \in (\widehat{\mathbb{Z}/q\mathbb{Z}})^\times$  such that  $\eta_1 \chi = \eta_2$  and  $\pi_1 \otimes \chi \simeq \pi_2$ , then we say that  $(\pi_1, \eta_1), (\pi_2, \eta_2) \in \mathcal{H}_{* \leq T}^{\text{tw}}(q)$  are *twist-equivalent* and write  $(\pi_1, \eta_1) \sim (\pi_2, \eta_2)$ . The relation

$\sim$  is an equivalence relation, and thus we may partition  $\mathcal{H}_{*\leq T}^{\text{tw}}(q)$  into twist classes  $\mathcal{T} \in \mathcal{H}_{*\leq T}^{\text{tw}}(q)/\sim$ . The twist classes also arise naturally as the fibers of the map

$$\Phi : \mathcal{H}_{*\leq T}^{\text{tw}}(q) \rightarrow \bigcup_{*\leq T} \mathcal{H}_*, \quad (\pi, \eta) \mapsto \pi \otimes \bar{\eta}.$$

With the notation defined in (7.8), we therefore have

$$(7.15) \quad \sum_{\eta \in (\mathbb{Z}/q\mathbb{Z})^\times} \int_{*\leq T} \sum_{m|q} \sum_{\pi \in \mathcal{H}_*(m, \eta^2)} |L(1/2, \pi \otimes \bar{\eta})|^4 = \sum_{\mathcal{T} \in \mathcal{H}_{*\leq T}^{\text{tw}}(q)/\sim} |L(1/2, \Phi(\mathcal{T}))|^4 |\mathcal{T}|,$$

where  $L(1/2, \Phi(\mathcal{T})) = L(1/2, \pi \otimes \bar{\eta})$  only depends on the twist class  $\mathcal{T}$ .

Note that the automorphic representation  $\Phi(\mathcal{T})$  has trivial central character and conductor dividing  $q^2$ . We now estimate the size of  $|\mathcal{T}|$  as well as its conductor in order to show that whenever the conductor of  $\Phi(\mathcal{T})$  is large, then  $|\mathcal{T}|$  is small to compensate, and vice-versa.

First of all, each twist class  $\mathcal{T}$  contains a pair  $(\pi, \eta)$  for which  $\pi$  is twist-minimal at all primes dividing  $q$ , and so we choose such a twist-minimal pair in each  $\mathcal{T}$ , say  $(\pi_{\mathcal{T}}, \eta_{\mathcal{T}})$ . By Lemma 6.2 (i.e. (6.6)) we have for any  $(\pi, \eta) \in \mathcal{T}$

$$(7.16) \quad \mathbf{c}(\pi \otimes \bar{\eta}) = \mathbf{c}(\pi_{\mathcal{T}} \otimes \bar{\eta}_{\mathcal{T}}) = [\mathbf{c}(\pi_{\mathcal{T}}), \mathbf{c}(\eta_{\mathcal{T}})^2].$$

Secondly, to estimate the sizes of the twist classes  $|\mathcal{T}|$  we have the following estimate.

**Lemma 7.7.** *For an integer  $n \geq 1$ , define  $\text{flrt}(n)$  to be the largest integer  $d$  so that  $d^2|n$ . For  $\mathcal{T} \in \mathcal{H}_{*\leq T}^{\text{tw}}(q)/\sim$  we have*

$$(7.17) \quad |\mathcal{T}| \ll \left( \text{flrt}(q), \frac{q}{\mathbf{c}(\eta_{\mathcal{T}}^2)} \right) q^\varepsilon.$$

*Proof.* Since every  $(\eta, \pi)$  in  $\mathcal{T}$  is a twist of  $(\eta_{\mathcal{T}}, \pi_{\mathcal{T}})$ , we have

$$(7.18) \quad |\mathcal{T}| = \#\{\chi \in (\mathbb{Z}/q\mathbb{Z})^\times : \mathbf{c}(\pi_{\mathcal{T}} \otimes \chi) \mid q\}.$$

By (6.6), the condition  $\mathbf{c}(\pi_{\mathcal{T}} \otimes \chi) \mid q$  is equivalent to  $\mathbf{c}(\chi) \mathbf{c}(\eta_{\mathcal{T}}^2 \chi) \mid q$ . Since we now see that  $|\mathcal{T}|$  is multiplicative in  $\eta$  and  $q$ , and since the bound (7.17) is also, it suffices to prove the lemma under the assumption that  $q$  is a prime power.

We now assume that  $q = p^v$ , and switch to the conductor exponent,  $c$ . For notational simplicity, replace  $\eta_{\mathcal{T}}$  by  $\eta$ . Note  $\text{flrt}(p^v) = p^{\lfloor v/2 \rfloor}$ . By the previous discussion, we have  $|\mathcal{T}| = \#\{\chi : c(\chi) + c(\eta^2 \chi) \leq v\}$ . There are three classes of  $\chi$  to consider:

- (1) The case  $c(\chi) \leq c(\eta^2)$  and  $c(\chi \eta^2) = c(\eta^2)$ , the latter condition being automatic if  $c(\chi) < c(\eta^2)$ . Under this assumption, the condition  $c(\chi) + c(\eta^2 \chi) \leq v$  is equivalent to  $c(\chi) + c(\eta^2) \leq v$ , which in turn implies  $c(\chi) \leq \min(c(\eta^2), v - c(\eta^2))$ . This last quantity is always  $\leq \lfloor v/2 \rfloor$ , so the number of characters  $\chi$  satisfying the above hypotheses is bounded as claimed in the lemma.
- (2) The case  $c(\chi) = c(\eta^2)$  and  $c(\chi \eta^2) < c(\eta^2)$ . Such  $\chi$  are of the form  $\chi = \bar{\eta}^2 \chi'$  with  $c(\chi') < c(\eta^2)$ . Then  $c(\chi) + c(\eta^2 \chi) = c(\chi') + c(\eta^2)$ , and so the number of  $\chi \in \mathcal{T}$  satisfying the hypotheses of this case is bounded as in the previous case.
- (3) The case  $c(\chi) > c(\eta^2)$ . This hypothesis implies that  $c(\chi \eta^2) = c(\chi)$ , and so  $c(\chi) + c(\eta^2 \chi) = 2c(\chi)$ . Thus any  $\chi$  in this case which satisfies  $c(\chi) + c(\eta^2 \chi) \leq v$  also has  $c(\chi) \leq \lfloor v/2 \rfloor$ . On the other hand, since  $c(\chi) > c(\eta^2)$  we also have  $v \geq c(\chi) + c(\eta^2 \chi) > c(\chi) + c(\eta^2)$ , so  $c(\chi) < v - c(\eta^2)$ , finishing the proof.  $\square$

Lastly, to control the set  $\mathcal{H}_{*\leq T}^{\text{tw}}(q)/\sim$  we will use that it is in bijection with the image  $\Phi(\mathcal{H}_{*\leq T}^{\text{tw}}(q))$  of  $\Phi$ . The conductors of the forms  $\sigma \in \Phi(\mathcal{H}_{*\leq T}^{\text{tw}}(q))$  are then given by (7.16).

The above three facts, along with the spectral large sieve inequality (Lemma 7.4) will suffice to finish the proof of Theorem 7.6. To implement them, we must parametrize the possible values of  $\mathfrak{c}(\eta_{\mathcal{T}})^2$ ,  $\mathfrak{c}(\eta_{\mathcal{T}}^2)$ , and  $\mathfrak{c}(\pi_{\mathcal{T}})$  that may occur as  $\mathcal{T}$  runs over  $\mathcal{H}_{*\leq T}^{\text{tw}}(q)/\sim$ . We thus write the right hand side of (7.15) as

$$(7.19) \quad \sum_{r|q} \sum'_{d|r} \sum_{\substack{m|q \\ d|m}} \sum_{\substack{\mathcal{T}:\mathfrak{c}(\eta_{\mathcal{T}})=r \\ \mathfrak{c}(\eta_{\mathcal{T}}^2)=d \\ \mathfrak{c}(\pi_{\mathcal{T}})=m}} |L(1/2, \Phi(\mathcal{T}))|^4 |\mathcal{T}|,$$

where the ' on the sum over  $d$  indicates that there are some extra constraints on the parameters  $r$  and  $d$ , which we now explicate for later use.

A character  $\eta \in (\widehat{\mathbb{Z}/q\mathbb{Z}})^{\times}$  of conductor  $r$  factors over places  $\eta = \prod_p \eta_p$ , where each  $\eta_p : \mathbb{Z}_p^{\times} \rightarrow \mathbb{C}^{\times}$  and  $\mathfrak{c}(\eta) = \prod_{p|r} p^{c_p(\eta_p)}$ . For  $p$  odd,  $c_p(\eta_p) = c_p(\eta_p^2)$  unless  $c_p(\eta_p) = 1$  and  $\eta_p$  is the Legendre symbol. When  $p = 2$ , one may similarly check that if  $c_2(\eta_2) = \beta$  with  $\beta \geq 4$  then  $c_2(\eta_2^2) = \beta - 1$ . Considering  $\beta = 2, 3$  separately, we can conclude that  $c_2(\eta_2) - c_2(\eta_2^2) \in \{1, 2, 3\}$ . Therefore, the ' on the sum in (7.19) indicates that the sum runs over those  $d | r$  such that there exists  $k \in \{1, 2, 3\}$  and an odd square-free integer  $r'$  such that  $r/d = 2^k r'$  and  $(d, r') = 1$ .

By positivity, Lemma 7.7, and (7.16) we have that the sum in (7.15) is

$$\ll q^{\varepsilon} \sum_{r|q} \sum'_{d|r} \sum_{\substack{m|q \\ d|m}} (\text{flrt}(q), \frac{q}{d}) \int_{*\leq T} \sum_{\sigma \in \mathcal{H}_*([m, r^2])} |L(1/2, \sigma)|^4.$$

Now we are in a position to apply the spectral large sieve Lemma 7.4 (more precisely, the special case  $q_2 = 1$  of Theorem 7.6), giving that (7.15) is

$$(7.20) \quad \ll T^2 (qT)^{\varepsilon} \sum_{r|q} \sum'_{d|r} \sum_{\substack{m|q \\ d|m}} (\text{flrt}(q), \frac{q}{d}) [m, r^2] \ll qT^2 (qT)^{\varepsilon} \sum_{r|q} \sum'_{d|r} \frac{(\text{flrt}(q)d, q)}{d} \frac{r^2}{(q, r^2)}.$$

Our goal is to show that this is  $\ll (qT)^{2+\varepsilon}$ , so it suffices to show that the innermost double sum in (7.20) is  $\ll q^{1+\varepsilon}$ , and since both sides are multiplicative, it suffices to show it when  $q$  is a prime power, which we now assume. If  $q$  is odd, then the conditions indicated by the ' imply that either  $d = r$ , or  $r$  is a prime and  $d = 1$ . In the case that  $d = r$ , observe that

$$r \frac{(\text{flrt}(q)r, q)}{(r^2, q)} = [\text{flrt}(q), r],$$

which implies the desired bound. The desired bound  $\ll q$  is even easier to check in the case that  $r$  is a prime and  $d = 1$ . If  $p = 2$  then one uses that  $\frac{r^2}{d} | 8r$  along with the previous reasoning to obtain the desired bound.  $\square$

## 8. HARMONIC ANALYSIS STEPS

We now begin the proof of Theorem 1.5. The sequence of steps used in the proof is motivated in Section 1.4. Let

$$S(\chi, h) = \sum_n \tau(n+h) \chi(n+h) \bar{\chi}(n) \sum_{n_1 n_2 = n} w(n_1, n_2, h),$$

with  $w(x, y, z)$  as in Section 1.3, which in particular satisfies (1.13), and has support on  $r \ll N$ . We also recall that  $\chi$  is primitive modulo  $q$ .

**8.1. Approximate functional equation.** Applying Lemma 7.3 to  $S(\chi, h)$ , we obtain

$$(8.1) \quad S(\chi, h) = \frac{2}{\tau(\bar{\chi})} \sum_c \frac{\chi(c)}{c} \sum_{r \pmod{cq}}^* \bar{\chi}(r) e_{cq}(hr) \sum_{n_1, n_2} e_{cq}(n_1 n_2 r) \bar{\chi}(n_1 n_2) w(n_1, n_2, h) f\left(\frac{c}{\sqrt{n_1 n_2 + h}}\right).$$

**8.2. Poisson summation.**

**Lemma 8.1.** *We have*

$$(8.2) \quad S(\chi, h) = \sum_{(c, q)=1} \frac{2}{c^2 q^2} \sum_{n_1, n_2 \in \mathbb{Z}} I(c, n_1, n_2, h) S(h\bar{q}, -n_1 n_2 \bar{q}; c) T_\chi(h, \bar{c} n_1, \bar{c} n_2),$$

where

$$(8.3) \quad I(y, t_1, t_2, h) = \int_0^\infty \int_0^\infty w(x_1, x_2, h) f\left(\frac{y}{\sqrt{x_1 x_2 + h}}\right) e_{yq}(-x_1 t_1 - x_2 t_2) dx_1 dx_2,$$

and

$$(8.4) \quad T_\chi(h, m, n) = \sum_{x, y \pmod{q}}^* \chi(x + h) \bar{\chi}(x) e_q(mx\bar{y} + ny).$$

Remark. By integration by parts, and using (7.4) it is easy to see that for  $r \ll N$ ,

$$(8.5) \quad \frac{\partial^{k+j_1+j_2+\ell}}{\partial y^k \partial t_1^{j_1} \partial t_2^{j_2} \partial r^\ell} I(y, t_1, t_2, r) \ll_{k, j_1, j_2, \ell, A_1, A_2, \varepsilon} \frac{N^{1+\varepsilon} y^{-k} |t_1|^{-j_1} |t_2|^{-j_2} r^{-\ell} y^{-\varepsilon} q^\varepsilon}{(1 + \frac{y^2}{N})^{A_1} (1 + \frac{|t_1| N_1}{yq})^{A_2} (1 + \frac{|t_2| N_2}{yq})^{A_2}},$$

where  $A_1, A_2 > 0$  may be taken to be arbitrarily large.

*Proof.* Consider a sum of the form

$$V = \sum_{n_1, n_2} e_{cq}(n_1 n_2 r) \bar{\chi}(n_1 n_2) W(n_1, n_2),$$

where  $W$  is smooth of compact support, and  $(r, cq) = (c, q) = 1$ . We split the sum into arithmetic progressions modulo  $cq$  and apply Poisson summation, giving

$$(8.6) \quad V = \frac{1}{c^2 q^2} \sum_{n_1, n_2} A(n_1, n_2, c, q, \chi) I(n_1, n_2),$$

where

$$A(n_1, n_2, c, q, \chi) = \sum_{x_1, x_2 \pmod{cq}} \bar{\chi}(x_1 x_2) e_{cq}(x_1 x_2 r + x_1 n_1 + x_2 n_2),$$

and

$$I(n_1, n_2) = \int_0^\infty \int_0^\infty W(x_1, x_2) e_{cq}(-x_1 n_1 - x_2 n_2) dx_1 dx_2.$$

By the Chinese remainder theorem, we have

$$A(n_1, n_2, c, q, \chi) = A_q(n_1, n_2, \chi) A_c(n_1, n_2),$$

where

$$A_q(n_1, n_2, \chi) = \sum_{x_1, x_2 \pmod{q}} \bar{\chi}(x_1 x_2) e_q(x_1 x_2 \bar{c} r + n_1 \bar{c} x_1 + n_2 \bar{c} x_2),$$

and

$$A_c(n_1, n_2) = \sum_{x_1, x_2 \pmod{c}} e_c(x_1 x_2 \bar{q} r + n_1 \bar{q} x_1 + n_2 \bar{q} x_2) = c e_c(-n_1 n_2 \bar{r} \bar{q}).$$

Now we insert into (8.1) the formula (8.6) and subsequent evaluations, obtaining

$$S(\chi, h) = \frac{2}{\tau(\bar{\chi})} \sum_c \frac{\chi(c)}{c^2 q^2} \sum_{r \pmod{cq}}^* \bar{\chi}(r) e_{cq}(hr) \sum_{n_1, n_2} A_q(n_1, n_2, \chi) e_c(-n_1 n_2 \bar{r} \bar{q}) I(c, n_1, n_2, h).$$

Next we evaluate the  $r$ -sum, that is

$$B(n_1, n_2, c, q, \chi) := \sum_{r \pmod{cq}}^* \bar{\chi}(r) e_{cq}(hr) A_q(n_1, n_2, \chi) e_c(-n_1 n_2 \bar{r} \bar{q}).$$

By the Chinese Remainder Theorem, we have

$$B(n_1, n_2, c, q, \chi) = B_c(n_1, n_2) B_q(n_1, n_2, \chi),$$

where

$$B_c(n_1, n_2) = \sum_{r \pmod{c}}^* e_c(hr \bar{q} - n_1 n_2 \bar{r} \bar{q}) = S(h \bar{q}, -n_1 n_2 \bar{q}; c),$$

and

$$B_q(n_1, n_2, \chi) = \sum_{r \pmod{q}}^* \bar{\chi}(r) e_q(hr \bar{c}) \sum_{x_1, x_2 \pmod{q}} \bar{\chi}(x_1 x_2) e_q(x_1 x_2 \bar{c} r + n_1 \bar{c} x_1 + n_2 \bar{c} x_2).$$

Evaluating the  $r$ -sum with Lemma 7.1, we obtain

$$B_q(n_1, n_2, \chi) = \tau(\bar{\chi}) \bar{\chi}(c) \sum_{x_1, x_2 \pmod{q}} \chi(x_1 x_2 + h) \bar{\chi}(x_1 x_2) e_q(n_1 \bar{c} x_1 + n_2 \bar{c} x_2).$$

By a simple change of variables, we have  $B_q(n_1, n_2, \chi) = \tau(\bar{\chi}) \bar{\chi}(c) T_\chi(h, \bar{c} n_1, \bar{c} n_2)$ , which completes the proof.  $\square$

**8.3. Estimation of zero frequency terms.** We begin with an elementary bound.

**Lemma 8.2.** *Suppose that  $M \geq 1$  and  $w = w_M$  satisfies*

$$x^j w^{(j)}(x) \ll_j \left(1 + \frac{x}{M}\right)^{-100}.$$

*Then for  $q \neq 1$  we have*

$$\sum_m w(m) S(m, 0; q) \ll (Mq)^\varepsilon \min(M, q).$$

*Proof.* By integration by parts, the derivative bounds on  $w$  imply that for  $\text{Re}(s) = \sigma \in (0, 2]$  the Mellin transform of  $w$  satisfies

$$\tilde{w}(s) = \int_0^\infty w(x) x^s \frac{dx}{x} \ll_\sigma M^\sigma |s(s+1) \dots (s+100)|^{-1}.$$

By Mellin inversion, we have

$$\sum_m w(m) S(m, 0; q) = \frac{1}{2\pi i} \int_{(2)} \tilde{w}(s) \left( \sum_{m=1}^\infty \frac{S(m, 0; q)}{m^s} \right) ds.$$

Note

$$\sum_{m=1}^{\infty} \frac{S(m, 0; q)}{m^s} = \zeta(s) \sum_{a|q} a^{1-s} \mu(q/a).$$

The crucial feature is that this Dirichlet series does not have a pole at  $s = 1$ , by the Möbius inversion formula, since  $q > 1$ . We may then freely move the contour of integration to either the  $\varepsilon$ -line or the  $(1 + \varepsilon)$ -line, leading to the claimed bound.  $\square$

**Lemma 8.3.** *We have*

$$(8.7) \quad S(\chi, h) = \sum_{(c,q)=1} \frac{2}{c^2 q^2} \sum_{n_1, n_2 \neq 0} I(c, n_1, n_2, h) S(h\bar{q}, -n_1 n_2 \bar{q}; c) T_\chi(h, \bar{c}n_1, \bar{c}n_2) + O\left(N \frac{(h, q)}{q} (qN)^\varepsilon\right).$$

Remark. This error term is consistent with (1.14), since

$$(8.8) \quad \sum_{\substack{1 \leq h \leq H \\ h \equiv 0 \pmod{d}}} \frac{(h, q)}{q} \ll \frac{H}{q} (Nq)^\varepsilon.$$

For simplicity, and since it suffices for our application of Theorem 1.5, we have bounded the main term trivially, but it could be extracted in explicit form with more work.

*Proof.* We need to bound the terms in (8.2) with some  $n_i = 0$ . Recall the bound on  $I$  from (8.5), and also observe that

$$T_\chi(h, 0, n) = S(n, 0; q) \sum_{x \pmod{q}}^* \chi(x+h) \bar{\chi}(x) = S(n, 0; q) S(h, 0; q),$$

where the  $x$ -sum may be evaluated by changing basis to additive characters. Therefore the contribution to  $S(\chi, h)$  from  $n_1 = n_2 = 0$  is

$$2 \sum_{(c,q)=1} \frac{\varphi(q)}{c^2 q^2} I(c, 0, 0, h) S(h, 0; c) S(h, 0; q) \ll N (Nq)^\varepsilon \frac{(h, q)}{q}.$$

Similarly, the contribution from  $n_1 = 0$  and  $n_2 \neq 0$  is

$$2 \sum_{(c,q)=1} \frac{S(h, 0; c) S(h, 0; q)}{c^2 q^2} \sum_{n_2 \neq 0} I(c, 0, n_2, h) S(n_2, 0; q).$$

Using Lemma 8.2 and (8.5), we deduce

$$\sum_{n_2 \neq 0} I(c, 0, n_2, h) S(n_2, 0; q) \ll (Nq)^\varepsilon N \min\left(\frac{cq}{N_2}, q\right).$$

Therefore the contribution to  $S(\chi, h)$  from  $n_1 = 0$  and  $n_2 \neq 0$  is

$$\ll N (Nq)^\varepsilon \frac{(h, q)}{q} \sum_c \frac{(h, c)}{c^2} \left(1 + \frac{c^2}{N}\right)^{-A} \min\left(1, \frac{c}{N_2}\right) \ll N_1 \frac{(h, q)}{q} (Nq)^\varepsilon.$$

Since  $N_1 \ll N$ , this is even better than the claimed bound that arose from  $n_1 = n_2 = 0$ . By a symmetry argument, a similar bound holds for the terms with  $n_2 = 0$  and  $n_1 \neq 0$ .  $\square$

At this point, we pause to record a crude bound for  $S(\chi, h)$ . Using the trivial bound  $|T_\chi(h, \bar{c}n_1, \bar{c}n_2)| \leq q^2$ , the Weil bound for Kloosterman sums, and the bound (8.5) on  $I$ , we deduce from Lemma 8.3:



**Corollary 8.4.** *We have*

$$S(\chi, h) \ll q^2 N^{3/4} (qN)^\varepsilon + N \frac{(h, q)}{q} (qN)^\varepsilon$$

Remarks. When  $q = 1$ , this corresponds to the classical  $N^{3/4+\varepsilon}$  error term for the smoothed shifted divisor problem, which follows from the Weil bound (e.g. see [DFI, Thm. 1]).

We emphasize that the arithmetical conditions on  $d, h, q$  appearing in the statement of Theorem 1.5 have not been used yet (though the archimedean condition  $h \ll N$  was used in (8.5)).

This bound is trivial for  $q \gg N^{1/8}$ , while the reader may recall that for Theorem 1.4, we need  $N \ll q$ . Naturally, our next aim will be to improve this bound by proving a non-trivial bound on  $T_\chi$ . An optimistic guess is that  $T_\chi$  is  $O(q^{1+\varepsilon})$  in a suitable average sense. If true, this would only improve the error term  $q^2 N^{3/4}$  to  $qN^{3/4}$ , which is still trivial for  $q \gg N^{1/4}$ , so more advanced techniques will be necessary.

**8.4. Properties of  $T_\chi$ .** Recall that  $T_\chi$  was defined by (8.4). First observe the symmetry

$$(8.9) \quad T_\chi(h, m, n) = T_\chi(h, n, m).$$

Suppose that  $\chi = \chi_1 \chi_2$  with  $\chi_j$  modulo  $q_j$  with  $(q_1, q_2) = 1$ . The Chinese remainder theorem gives

$$(8.10) \quad T_\chi(h, m, n) = T_{\chi_1}(h, m\bar{q}_2, n\bar{q}_2) T_{\chi_2}(h, m\bar{q}_1, n\bar{q}_1).$$

So far the conditions  $d|q$ ,  $q^3|d^2$ , and  $h \equiv 0 \pmod{d}$  appearing in Theorem 1.5 have not been used. In the next lemma, we give a simplification of  $T_\chi$  under the assumption  $q|d^2$  (whence  $q|h^2$ ), which is weaker than the condition  $q^2|d^3$  (whence  $q^2|h^3$ ) that will be used later. We write

$$h_q = (h, q).$$

**Lemma 8.5.** *Suppose  $q|h^2$ . If  $q|h$  then*

$$(8.11) \quad T_\chi(h, m, n) = S(m, 0; q) S(n, 0; q).$$

*If  $v_p(h) < v_p(q)$  for each  $p|q$ , then  $T_\chi(h, m, n) = 0$  unless  $h_q|(m, n)$ . In that case, we have*

$$(8.12) \quad T_\chi(h, m, n) = h_q^2 \text{Kl}_3\left(\ell_\chi \frac{h}{h_q}, \frac{m}{h_q}, \frac{n}{h_q}; \frac{q}{h_q}\right),$$

*where  $\text{Kl}_3(a, b, c; q)$  is the hyper-Kloosterman sum defined by*

$$\text{Kl}_3(a, b, c; q) = \sum_{\substack{x, y, z \pmod{q} \\ xyz \equiv 1 \pmod{q}}} e_q(ax + by + cz),$$

*and for  $\ell_\chi$  a certain integer with  $(\ell_\chi, q) = 1$ .*

Remarks.

- The two cases considered in Lemma 8.5, along with (8.10), are enough to completely evaluate  $T_\chi$  (see Corollary 8.6 below).
- The integer  $\ell_\chi$  appearing in Lemma 8.5 is such that when  $q | h_q^2$  we have  $\chi(1 + h_q t) = e_{q/h_q}(\ell_\chi t)$  for all  $t \in \mathbb{Z}$ . Warning: there is some subtlety involved in passing from the locally defined  $\ell_\chi$  appearing in Lemma 2.1 to the global  $\ell_\chi$  defined here.

*Proof.* If  $q|h$ , then (8.11) is immediate from (8.4).

Next suppose that  $1 \leq v_p(h) < v_p(q)$  for all  $p|q$ . Since  $q|h^2$ , we have  $\chi(1+hu)\chi(1+hv) = \chi(1+h(u+v))$ , which means  $\chi(1+hu)$ , as a function of  $u$ , is an additive character modulo  $q/h_q$ . Therefore, there exists an integer  $\ell_\chi$  so that  $\chi(1+hu) = e_{q/h_q}(\ell_\chi h'u)$ , where  $h = h'h_q$ . Since  $\chi$  is primitive, and  $q/h_q$  shares the same set of prime factors as  $q$ , then  $(\ell_\chi, q) = 1$ . In particular,  $\chi(x+h)\overline{\chi}(x) = \chi(1+h\overline{x}) = e_{q/h_q}(\ell_\chi h'\overline{x})$ , and so

$$T_\chi(h_q h', m, n) = \sum_{x \pmod{q}}^* \sum_{y \pmod{q}}^* e_{q/h_q}(\ell_\chi h'\overline{x}) e_q(mx\overline{y} + ny).$$

Changing variables  $x \rightarrow x + \frac{q}{h_q}$  shows the sum vanishes unless  $h_q|m$ , in which case it is the same sum repeated  $h_q$  times. The same argument with  $y$  shows it vanishes unless  $h_q|n$ , and then gives (8.12).  $\square$

**Corollary 8.6.** *Let notation and assumptions be as in Lemma 8.5. Write  $q = q_1 q_2$  where*

$$(8.13) \quad q_1 = \prod_{\substack{p|q \\ v_p(q) \leq v_p(h)}} p^{v_p(q)}, \quad q_2 = \prod_{\substack{p|q \\ v_p(q) > v_p(h)}} p^{v_p(q)},$$

so that  $(q_1, q_2) = 1$ . Write  $\chi = \chi_1 \chi_2$  where  $\chi_j$  has conductor  $q_j$ . Let

$$h_{q_2} = (h, q_2), \quad \text{so} \quad (h, q) = q_1 h_{q_2}.$$

Then  $T_\chi(h, \overline{c}n_1, \overline{c}n_2) = 0$  unless  $h_{q_2} | (n_1, n_2)$ , in which case

$$T_\chi(h_{q_2} h', \overline{c}h_{q_2} n_1, \overline{c}h_{q_2} n_2) = h_{q_2}^2 S(0, n_1; q_1) S(0, n_2; q_1) \text{Kl}_3\left(\ell_{\chi_2} h', \overline{c}q_1 n_1, \overline{c}q_1 n_2; \frac{q_2}{h_{q_2}}\right).$$

*Remark.* The factorization of  $q$  as  $q_1 q_2$  is very natural, since the  $q_1$ -part of  $\chi$  is almost irrelevant, since  $\chi_1(n+h)\overline{\chi}_1(n) = 1$  for  $(n, q_1) = 1$ .

Applying Corollary 8.6 to (8.7), with  $h = h'h_{q_2}$  and with the replacements  $n_i \mapsto h_{q_2} n_i$ , we deduce

$$(8.14) \quad S(\chi, h) = \left[ \sum_{(c,q)=1} \frac{2h_{q_2}^2}{c^2 q^2} \sum_{n_1, n_2 \neq 0} I(c, h_{q_2} n_1, h_{q_2} n_2, h' h_{q_2}) S(h_{q_2} h' \overline{q}, -n_1 n_2 h_{q_2}^2 \overline{q}; c) \right. \\ \left. S(0, n_1; q_1) S(0, n_2; q_1) \text{Kl}_3\left(\ell_{\chi_2} h', \overline{c}q_1 n_1, \overline{c}q_1 n_2; \frac{q_2}{h_{q_2}}\right) \right] + O\left(N \frac{(h, q)}{q} (qN)^\varepsilon\right).$$

Again, we pause our analysis to record a simple bound on  $S(\chi, h)$ .

**Corollary 8.7.** *Suppose  $d|(h, q)$  with  $q|d^2$  and write  $q = q_1 q_2$  as in (8.13). Then*

$$S(\chi, h) \ll N^{3/4} \frac{q_2}{h_{q_2}} (qN)^\varepsilon + N \frac{(h, q)}{q} (qN)^\varepsilon,$$

and

$$(8.15) \quad S(\chi) \ll (N^{3/4} \frac{Hq}{d^2} + N \frac{H}{q}) (qN)^\varepsilon.$$

*Proof.* Smith [Sm, Thm. 6] showed  $\text{Kl}_3(a, b, c; q) \ll q^{1+\varepsilon}$  assuming  $(a, q) = 1$  (of course, the most difficult case where  $q$  is prime and  $(abc, q) = 1$  is due to Deligne [De2, Sommes trig. §7]). Applying this bound to (8.14), noting that  $(\ell_{\chi_2} h', \frac{q_2}{h_{q_2}}) = 1$ , easily finishes the proof.  $\square$

Remarks.

- Observe the significant savings of the factor  $\frac{q_2}{h_{q_2}}$  compared to the optimistic factor of  $q$  in the discussion following Corollary 8.4. This arose in large part from the important feature that  $h_{q_2} | (n_1, n_2)$  (which essentially cancelled against the  $h_{q_2}^2$  factor in (8.14)) while the modulus of  $T_\chi$  is greatly reduced to  $q_2/h_{q_2}$ . We also save a  $q_1$ -factor from the boundedness (on average) of Ramanujan sums.
- The bound (8.15) is consistent with our ultimate goal (1.14) when  $N \asymp q$  provided that  $d \gg q^{7/8}$ , which is still a bit too restrictive, since the condition  $q^2 | d^3$  means  $d$  can be as small as  $q^{2/3}$ .
- The bound (8.15) is compatible with the discussion in the sketch following (1.18).

### 8.5. Some properties of $\text{Kl}_3$ , and application to $S(\chi, h)$ .

**Lemma 8.8.** *Suppose that  $q \geq 1$  and  $a, b, c \in \mathbb{Z}$ . Write  $a = a_0 a'$ ,  $b = b_0 b'$ ,  $c = c_0 c'$  where  $a_0 b_0 c_0 | q^\infty$  and  $(a' b' c', q) = 1$ . Then*

$$\text{Kl}_3(a, b, c; q) = \frac{1}{\varphi(q)} \sum_{\eta \pmod{q}} \tau(\bar{\eta}, a_0) \tau(\bar{\eta}, b_0) \tau(\bar{\eta}, c_0) \eta(a' b' c'),$$

where for  $n \in \mathbb{Z}$  the Gauss sum  $\tau(\eta, n)$  was defined in Lemma 7.1.

*Proof.* The proof is standard, a similar calculation appearing in e.g. [Sm, §2].  $\square$

We apply Lemma 8.8 to (8.14). We already saw in the proof of Corollary 8.7 that  $(\ell_{\chi_2} h', \frac{q_2}{h_{q_2}}) = 1$ . To handle  $n_1, n_2$ , we write  $n_1 = n_{10} n'_1$  and  $n_2 = n_{20} n'_2$  where  $n_{10} n_{20} | q_2^\infty$  and  $(n'_1 n'_2, q_2) = 1$ . These two conditions are equivalent to  $(n'_1 n'_2, q_2/h_{q_2}) = 1$  and  $n_{10} n_{20} | (q_2/h_{q_2})^\infty$  since by definition (8.13) of  $q_2$ , the numbers  $q_2$  and  $q_2/h_{q_2}$  have the same set of prime factors. Thus

$$\begin{aligned} \text{Kl}_3\left(\ell_{\chi_2} h', \overline{c q_1} n_1, \overline{c q_1} n_2; \frac{q_2}{h_{q_2}}\right) \\ = \frac{1}{\varphi(q_2/h_{q_2})} \sum_{\eta \pmod{q_2/h_{q_2}}} \bar{\eta}^2(q_1) \tau(\bar{\eta}) \tau(\bar{\eta}, n_{10}) \tau(\bar{\eta}, n_{20}) \bar{\eta}^2(c) \eta(\ell_{\chi_2} h' n'_1 n'_2). \end{aligned}$$

Therefore,

$$(8.16) \quad S(\chi, h) = \left( \frac{2h_{q_2}^2}{q^2 \varphi(q_2/h_{q_2})} \sum_{\eta \pmod{q_2/h_{q_2}}} \bar{\eta}^2(q_1) \eta(\ell_{\chi_2} h') \sum_{n_{10}, n_{20} | q_2^\infty} \tau(\bar{\eta}) \tau(\bar{\eta}, n_{10}) \tau(\bar{\eta}, n_{20}) \right. \\ \left. \sum_{n'_1, n'_2 \neq 0} S(0, n'_1; q_1) S(0, n'_2; q_1) \eta(n'_1 n'_2) K \right) + O\left(N \frac{(h, q)}{q} (qN)^\varepsilon\right),$$

where  $h = h' h_{q_2}$  and  $K$  is shorthand for the following sum of Kloosterman sums:

$$(8.17) \quad K = \sum_{(c, q)=1} \frac{I(c, h_{q_2} n_1, h_{q_2} n_2, h' h_{q_2})}{c^2} S(h_{q_2} h' \bar{q}, -n_1 n_2 h_{q_2}^2 \bar{q}; c) \bar{\eta}^2(c),$$

with  $n_i = n_{i0} n'_i$ .

## 9. SPECTRAL ANALYSIS OF THE SHIFTED DIVISOR SUM

**9.1. Set-up for Bruggeman-Kuznetsov.** Our next major goal is to apply the Bruggeman-Kuznetsov formula to  $K$  defined by (8.17). We begin with some simplifications. The evaluation of  $T_\chi$  appearing in Lemma 8.5 used that  $q|d^2$ , which is a weaker condition than what is assumed in Theorem 1.5, namely that  $q^2|d^3$ . We will use this stronger assumption now.

Recall that we factor  $q = q_1 q_2$  according to the value of  $h$  as in Corollary 8.6, and correspondingly set  $h_{q_2} = (h, q_2)$  and  $h = h_{q_2} h'$ . By the definition of the factorization  $q = q_1 q_2$ , this implies that  $q_1|h'$ . Recalling that  $d|h$ , for  $p|q_2$  we have  $3v_p(h_{q_2}) \geq 2v_p(q_2)$ , and therefore  $h_{q_2}^2/q_2$  is an integer. Note the hypothesis  $(c, q) = 1$  from (8.17) and observe

$$S(h_{q_2} h' \bar{q}, -n_1 n_2 h_{q_2}^2 \bar{q}; c) = S\left(\frac{h'}{q_1}, -n_1 n_2 \frac{h_{q_2}^2}{q_2} \overline{q_1 (q_2/h_{q_2})}; c\right),$$

using that  $S(ax, y; c) = S(x, ay; c)$  for  $(a, c) = 1$ .

Another small remark is that the condition  $(c, q) = 1$  in the definition of  $K$  is equivalent to the pair of conditions  $(c, q_1) = 1$  and  $(c, q_2) = 1$ . The latter of these is enforced by the presence of  $\bar{\eta}^2(c)$ , since  $q_2/h_{q_2}$  has the same prime factors as  $q_2$  (by definition of  $q_2$ ). The condition  $(c, q_1) = 1$  can be detected by multiplying (8.17) by the trivial character modulo  $q_1$ , which we denote  $\chi_{0, q_1}$ .

Taking these observations together, we derive that  $K$  equals

$$(9.1) \quad \sum_{(c, q)=1} \frac{I(c, h_{q_2} n_1, h_{q_2} n_2, h' h_{q_2})}{c^2} S\left(\frac{h'}{q_1}, -n_1 n_2 \frac{h_{q_2}^2}{q_2} \overline{q_1 (q_2/h_{q_2})}; c\right) \bar{\eta}^2(c) \chi_{0, q_1}(c).$$

Our next goal is to see that  $K$  can be viewed as an instance of  $\mathcal{K}$  (as in (6.24)), with the following choices of parameters. The level, say  $r$ , is given by

$$r = q_1 \frac{q_2}{h_{q_2}} = \frac{q}{h_{q_2}} = \frac{q}{(h, q_2)},$$

the central character is  $\eta^2 \chi_{0, q_1}$ , the pair of cusps is  $\infty, 0$  and  $m = h'/q_1$ ,  $n = -n_1 n_2 h_{q_2}^2/q_2$ . Define the weight function  $\Phi(y) = \Phi(y, \cdot)$  (we temporarily suppress the dependence of  $\Phi$  on the other variables) by

$$(9.2) \quad \Phi(y) = \frac{r}{y^2} I(r^{-1/2} y, t_1, t_2, h), \quad \text{with} \quad t_i = h_{q_2} n_i.$$

Therefore,  $K = \mathcal{K}$  as in (6.24), and hence by Theorem 6.10, we have  $K = K_{\text{Maass}} + K_{\text{Eis}} + K_{\text{hol}}$ , where

$$(9.3) \quad K_{\text{Maass}} = \sum_{t_j} \mathcal{L}^\pm \Phi(t_j) \sum_{\ell m = \frac{q}{h_{q_2}}} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \frac{4\pi \epsilon_\pi^{(\pm)}}{V(q/h_{q_2}) \mathcal{L}_\pi^*(1)} \sum_{\delta|\ell} \bar{\lambda}_\pi^{(\delta)}\left(\frac{h'}{q_1}\right) \bar{\lambda}_\pi^{(\delta)}\left(n_{10} n_{20} |n'_1 n'_2| \frac{h_{q_2}^2}{q_2}\right).$$

and similar formulas hold for  $K_{\text{Eis}}$  and  $K_{\text{hol}}$ .

Inserting this into (8.16), we correspondingly write

$$(9.4) \quad S(\chi, h) = S_{\text{Maass}}(\chi, h) + S_{\text{Eis}}(\chi, h) + S_{\text{hol}}(\chi, h) + O\left(N \frac{(h, q)}{q} (qN)^\varepsilon\right).$$

**9.2. The behavior of the integral transforms.** Here we study the analytic behavior of the integral transforms  $\mathcal{L}^\pm \Phi$  that occur in the Bruggeman-Kuznetsov formula.

**Lemma 9.1.** *For each  $q \in \mathbb{N}$ , let  $I = I(y, t_1, t_2, h)$  be a smooth function on  $\mathbb{R}_{>0} \times \mathbb{R}_{\neq 0}^2 \times \mathbb{R}_{>0}$  supported on  $h \ll H \ll N$  in the last variable, and satisfying the bounds (8.5) with  $N_1 N_2 = N$ . Let  $\Phi$  be defined in terms of  $I$  by (9.2), which is a function of the variables  $y, q, h, n_1, n_2$  via the substitutions  $r = \frac{q}{h_{q_2}}$  and  $t_i = h_{q_2} n_i$ ,  $i = 1, 2$ . Let us write*

$$|mn| := \frac{h'}{q_1} |n_1 n_2| \frac{h_{q_2}^2}{q_2} = \frac{h}{q} |n_1 n_2| h_{q_2}.$$

Then, there exists a function  $H = H_\pm(s, t, t_1, t_2, h)$  so that

$$(9.5) \quad \mathcal{L}^\pm \Phi(t) = \int_{(\sigma)} H_\pm(s, t, t_1, t_2, h) (mn)^{-s/2} ds,$$

where  $H_\pm$  is holomorphic in  $s$  for  $\operatorname{Re}(s/2) > \theta$ , with  $\theta$  any bound towards the Ramanujan conjecture. If  $|t| \gg (Nq|n_1 n_2|)^\varepsilon$  then

$$(9.6) \quad \mathcal{L}^\pm \Phi(t) \ll_{A, \varepsilon} (1 + |t|)^{-A} (Nq|n_1 n_2|)^{-100}.$$

If  $|t| \ll (qN|n_1 n_2|)^\varepsilon$ , then for  $\operatorname{Re}(s) > 1$  it satisfies the bound

$$(9.7) \quad H_\pm(s, t, t_1, t_2, h) \ll_\sigma (qN)^\varepsilon (Nr)^{\frac{1}{2} + \frac{\sigma}{2}} (1 + |s|)^{-A} \left(1 + \frac{|t_1| N_1}{\sqrt{Nq}}\right)^{-A} \left(1 + \frac{|t_2| N_2}{\sqrt{Nq}}\right)^{-A}.$$

Similarly, there exists a function  $H = H_{\text{hol}}(s, k, t_1, t_2, h)$  that is holomorphic in  $s$  for  $\operatorname{Re}(s) > 0$  so that

$$\mathcal{L}^{\text{hol}} \Phi(k) = \int_{(\sigma)} H_{\text{hol}}(s, k, t_1, t_2, h) (mn)^{-s/2} ds,$$

and satisfies the same bound as (9.7) for  $\operatorname{Re}(s) > 1$  when  $k \ll (qN|n_1 n_2|)^\varepsilon$ . If  $k \gg (Nq|n_1 n_2|)^\varepsilon$  then

$$\mathcal{L}^{\text{hol}} \Phi(k) \ll_{A, \varepsilon} k^{-A} (qN|n_1 n_2|)^{-100}.$$

Remarks. Since the claimed bounds are the same for the three choices of  $\pm, \text{hol}$ , we may easily treat these cases in unison. In addition, the assumption that  $t_1, t_2 \neq 0$  is unobjectionable by the reduction in Lemma 8.3.

*Proof.* We focus on the case of  $\mathcal{L}^\pm$  first. Recall that  $\mathcal{L}^\pm \Phi(t)$  is given by (6.15), where

$$\tilde{\Phi}(s+1) = r \int_0^\infty I(r^{-1/2} y, t_1, t_2, h) y^{s-2} dy = r^{\frac{s+1}{2}} \int_0^\infty I(y, t_1, t_2, h) y^{s-1} \frac{dy}{y}.$$

By (8.5), and using that  $t_1 t_2 \neq 0$ , the function  $I$  has rapid decay at 0 and  $\infty$ , and so  $\tilde{\Phi}$  is entire. One sees that  $\mathcal{L}^\pm \Phi$  has an integral representation of the form (9.5), with

$$(9.8) \quad H_\pm(s, t, t_1, t_2, h) = \frac{(4\pi)^{-s}}{2\pi i} \tilde{\Phi}(s+1) h_\pm(s, t),$$

where recall  $h_\pm(s, t)$  is defined by (6.16). It is easy to check that  $h_\pm(s, t)$  is analytic for  $\operatorname{Re}(s/2) > \theta$  since  $t \in \mathbb{R}$  or  $-\theta \leq it \leq \theta$ .

Next we work out bounds on  $\tilde{\Phi}$ . The decay of  $I(y)$  from (8.5) means in practice that  $y \ll \sqrt{N}$  and  $y \gg \max(\frac{|t_1| N_1}{q}, \frac{|t_2| N_2}{q})$ . For any  $a, b > 0$ , and  $s \neq 1$  with  $\operatorname{Re}(s) = \sigma$  we have

$|\int_a^b y^{s-2} dy| \leq |s-1|^{-1}(b^{\sigma-1} + a^{\sigma-1})$ . Integrating by parts, using the triangle inequality, and then using this bound, we deduce for any  $j \geq 0$  that

$$(9.9) \quad \tilde{\Phi}(s+1) \ll_{j,\sigma,A} \frac{Nr^{\frac{1}{2}+\frac{\sigma}{2}}(Nq)^\varepsilon}{|(s-1)\dots(s+j-2)|} \frac{N^{\frac{\sigma-1}{2}} + \max\left(\frac{|t_1|N_1}{q}, \frac{|t_2|N_2}{q}\right)^{\sigma-1}}{\left(1 + \frac{|t_1|N_1}{\sqrt{Nq}}\right)^A \left(1 + \frac{|t_2|N_2}{\sqrt{Nq}}\right)^A}.$$

For future use, we also record derivative bounds with respect to the other variables, which leads to the following minor generalization of (9.9), valid for  $\text{Re}(s) = \sigma > 1$ :

$$(9.10) \quad |t_1|^{j_1}|t_2|^{j_2}h^{j_3} \frac{\partial^{j_1+j_2+j_3}}{\partial t_1^{j_1} \partial t_2^{j_2} \partial h^{j_3}} \tilde{\Phi}(s+1, t_1, t_2, h) \\ \ll_{j,\sigma,A} \frac{Nr^{\frac{1}{2}+\frac{\sigma}{2}}(Nq)^\varepsilon N^{\frac{\sigma-1}{2}} + \max\left(\frac{|t_1|N_1}{q}, \frac{|t_2|N_2}{q}\right)^{\sigma-1}}{(1+|s|)^A \left(1 + \frac{|t_1|N_1}{\sqrt{Nq}}\right)^A \left(1 + \frac{|t_2|N_2}{\sqrt{Nq}}\right)^A}.$$

In addition, it is supported on  $h \ll H$ .

We also need a bound on  $h_\pm(s, t)$ . Let  $d(\cdot, \cdot)$  be the distance function on  $\mathbb{C}$ . Stirling's approximation shows that for  $\sigma$  fixed and with  $d(\frac{\sigma}{2}, \mathbb{Z}_{\leq 0}) \geq \frac{1}{100}$  we have

$$(9.11) \quad h_\pm(\sigma + iv, t) \ll (1 + |t + \frac{v}{2}|)^{\frac{\sigma-1}{2}} (1 + |t - \frac{v}{2}|)^{\frac{\sigma-1}{2}}.$$

Now it is easy to derive the bound (9.7) by putting together (9.8), (9.9), and (9.11).

It remains to show (9.6). To see this, suppose  $|t| \gg (Nq|n_1n_2|)^\varepsilon$  and shift the contour of integration in (6.15) far to the left. There are poles of  $h_\pm(s, t)$  at  $s/2 \pm it = 0, -1, -2, \dots$ , which have  $|\text{Im}(s)| \asymp |t| \gg (Nq|n_1n_2|)^\varepsilon$ . Since  $\tilde{\Phi}(s+1)$  is small at this height, these residues give a contribution to  $\mathcal{L}^\pm \Phi(t)$  that are consistent with (9.6). From a trivial bound on the new line  $\sigma$  with  $d(\frac{\sigma}{2}, \mathbb{Z}_{\leq 0}) \geq \frac{1}{100}$ , we obtain

$$\mathcal{L}^\pm \Phi(t) \ll \frac{Nr}{\sqrt{mn}} \left(\frac{|t|\sqrt{ra}}{\sqrt{mn}}\right)^{\sigma-1} + O(t^{-A}(Nq|n_1n_2|)^{-100}),$$

with

$$a = \max\left(\frac{|t_1|N_1}{q}, \frac{|t_2|N_2}{q}\right).$$

Here  $a$  is temporary shorthand notation not used past (9.12) below.

Note that

$$(9.12) \quad \frac{\sqrt{mn}}{a\sqrt{r}} = \frac{\sqrt{|n_1n_2|}\sqrt{h}}{\max(|n_1|N_1, |n_2|N_2)} \ll \frac{\sqrt{n_1n_2}\sqrt{N_1N_2}}{\max(|n_1|N_1, |n_2|N_2)},$$

where in the last bound we used  $h \ll N = N_1N_2$ . If  $|t| \gg (qN|n_1n_2|)^\varepsilon$ , we may take  $\sigma$  far to the left to see that  $\mathcal{L}^\pm \Phi(t)$  is very small, as desired.

Now we quickly treat the holomorphic case. If  $s = \sigma + iv$  with  $\sigma$  fixed and  $d(\frac{k+\sigma-1}{2}, \mathbb{Z}_{\leq 0}) \geq \frac{1}{100}$ , then analogously to (9.11), we have

$$(9.13) \quad \left|2^{s-1} \frac{\Gamma(\frac{k}{2} + \frac{s-1}{2})}{\Gamma(\frac{k}{2} - \frac{s-1}{2})}\right| \ll |k + iv|^{\sigma-1}.$$

An essentially identical method now shows that  $\mathcal{L}^{\text{hol}} \Phi(k)$  is very small if  $k \gg (qN|n_1n_2|)^\varepsilon$ .  $\square$

**9.3. The spectral expansion of  $S(\chi, h)$ .** We will mainly focus on  $S_{\text{Maass}}(\chi, h)$ .

We apply Lemma 9.1 to the function  $I$  defined in (8.3), and insert the result in (9.3) to derive (with  $t_i = h_{q_2} n_i$ ,  $n_i = n_{i0} n'_i$ )

$$(9.14) \quad K_{\text{Maass}} = \sum_{t_j} \int_{(2+\varepsilon)} H_{\pm}(s, t_j, t_1, t_2, h_{q_2} h') \left( \frac{q}{h_{q_2}^2 h' n_{10} n_{20} |n'_1 n'_2|} \right)^{s/2} ds \\ \times \sum_{\ell m = \frac{q}{h_{q_2}}} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \frac{4\pi \epsilon_{\pi}^{(\pm)}}{V(q/h_{q_2}) \mathcal{L}_{\pi}^*(1)} \sum_{\delta|\ell} \bar{\lambda}_{\pi}^{(\delta)} \left( \frac{h'}{q_1} \right) \bar{\lambda}_{\pi}^{(\delta)} \left( n_{10} n_{20} |n'_1 n'_2| \frac{h_{q_2}^2}{q_2} \right).$$

Looking back to (8.16), our next goal is to convert the sums over  $n'_1, n'_2$  to integrals of Dirichlet series, and to do so we must reduce to positive values of  $n'_i$ . We collect the contributions from the four quadrants of the  $n'_1, n'_2$  summation by setting for  $t_1, t_2 > 0$

$$(9.15) \quad H_1(s, t, t_1, t_2, h) = H_+(s, t, t_1, t_2, h) + \eta(-1) H_-(s, t, -t_1, t_2, h) \\ + \eta(-1) H_-(s, t, t_1, -t_2, h) + H_+(s, t, -t_1, -t_2, h).$$

By Mellin inversion there exists a function  $H_2$  such that

$$(9.16) \quad H_1(s, t, t_1, t_2, h) = \int_{\mathbf{u}} t_1^{-u_1} t_2^{-u_2} h^{-u_3} H_2(s, t, u_1, u_2, u_3) d\mathbf{u},$$

where we write  $\mathbf{u} = (u_1, u_2, u_3)$ , and which is absolutely convergent for  $\text{Re}(u_j) > 0$  for all  $j$ . Using (9.10), we have for  $\text{Re}(s) > 1$ ,  $\text{Re}(u_j) > 0$ , and  $t_j \ll (Nq)^{\varepsilon}$

$$(9.17) \quad H_2(s, t_j, u_1, u_2, u_3) \ll (Nr)^{\frac{1+\sigma}{2}} (Nq)^{\varepsilon} M_1^{\alpha_1} M_2^{\alpha_2} H^{\alpha_3} (1 + |s|)^{-A} \prod_{j=1}^3 (1 + |u_j|)^{-A},$$

where  $\alpha_j = \text{Re}(u_j)$ , and

$$(9.18) \quad M_j = \frac{\sqrt{Nq}}{N_j}, \quad j = 1, 2.$$

Gathering together (8.16), (9.4), (9.14), (9.15), and (9.16) we get

$$(9.19) \quad S_{\text{Maass}}(\chi, h_{q_2} h') = \frac{2h_{q_2}^2}{q^2 \varphi(q_2/h_{q_2})} \sum_{\eta \pmod{q_2/h_{q_2}}} \bar{\eta}^2(q_1) \eta(\ell_{\chi_2} h') \sum_{t_j} \sum_{\ell m = \frac{q}{h_{q_2}}} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \\ \frac{4\pi \epsilon_{\pi}^{(\pm)}}{V(q/h_{q_2}) \mathcal{L}_{\pi}^*(1)} \int_{\mathbf{u}} \int_{(2+\varepsilon)} \left( \frac{q}{h_{q_2}^2} \right)^{s/2} \frac{H_2(s, u_1, u_2, u_3)}{h_{q_2}^{u_1+u_2+u_3} (h')^{\frac{s}{2}+u_3}} Z(s, u_1, u_2; h_{q_2} h') ds d\mathbf{u},$$

where

$$(9.20) \quad Z(s, u_1, u_2; h_{q_2} h') = \sum_{\delta|\ell} \sum_{n_{10} n_{20} |q_2^{\infty}} \frac{\tau(\bar{\eta}) \tau(\bar{\eta}, n_{10}) \tau(\bar{\eta}, n_{20})}{n_{10}^{\frac{s}{2}+u_1} n_{20}^{\frac{s}{2}+u_2}} \\ \sum_{n'_1, n'_2 \geq 1} \frac{S(0, n'_1; q_1) S(0, n'_2; q_1)}{(n'_1)^{\frac{s}{2}+u_1} (n'_2)^{\frac{s}{2}+u_2}} \eta(n'_1 n'_2) \bar{\lambda}_{\pi}^{(\delta)} \left( \frac{h'}{q_1} \right) \bar{\lambda}_{\pi}^{(\delta)} \left( n_{10} n_{20} n'_1 n'_2 \frac{h_{q_2}^2}{q_2} \right).$$

For clarity, we recollect the origin of the relevant variable names. Firstly,  $q$  and  $d$  are given integers with  $d|q$ ,  $q^2|d^3$ . For Theorem 1.5, we want to sum over  $h \equiv 0 \pmod{d}$ . In our

analysis so far, without summation over  $h$ , we factored  $q$  according to  $h$ , via  $q = q_1 q_2$  where  $q_1 | h$  and  $1 \leq v_p(h) < v_p(q_2)$  for all  $p | q_2$ .

Our next task is to sum (9.19) over  $h = h_{q_2} h' \equiv 0 \pmod{d}$  and move the sum over  $h'$  to the inside. To implement this swap of the order of summation, we parametrize over all factorizations  $q = q_1 q_2$  with  $(q_1, q_2) = 1$  and all possible values of  $h_{q_2}$  satisfying  $h_{q_2} | q_2$  with  $v_p(h_{q_2}) < v_p(q_2)$  for all  $p | h_{q_2}$ . Note that these constraints enforce  $(h, q_2) = h_{q_2}$ , and  $h \equiv 0 \pmod{q_1}$ . We may also factor  $d = d_1 d_2$  where  $d_1 | q_1$  and  $d_2 | q_2$ , and then only sum over  $h_{q_2}$  with  $d_2 | h_{q_2}$ . In this way, we obtain

$$(9.21) \quad \sum_{h \equiv 0 \pmod{d}} S_{\text{Maass}}(\chi, h) = \sum_{\substack{q_1 q_2 = q \\ (q_1, q_2) = 1}} \sum_{\substack{h_{q_2} | q_2 \\ v_p(h_{q_2}) < v_p(q_2) \\ d_2 | h_{q_2}}} \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h').$$

Caution: the status of  $h_{q_2}$  has changed. Prior to (9.21),  $h_{q_2}$  was a function of  $q$  and  $h$ , whereas now it is only a summation variable constrained by the conditions indicated above. Thus,

$$(9.22) \quad \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h') = \frac{2h_{q_2}^2}{q^2 \varphi(q_2/h_{q_2})} \sum_{\eta \pmod{q_2/h_{q_2}}} \bar{\eta}^2(q_1) \eta(\ell_{\chi_2}) \sum_{t_j} \sum_{\ell m = \frac{q}{h_{q_2}}} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \\ \frac{4\pi \epsilon_\pi^{(\pm)}}{V(q/h_{q_2}) \mathcal{L}_\pi^*(1)} \int_{\mathbf{u}} \int_{(2+\varepsilon)} \left(\frac{q}{h_{q_2}^2}\right)^{s/2} \frac{H_2(s, u_1, u_2, u_3)}{h_{q_2}^{u_1+u_2+u_3}} \mathcal{Z}(s, u_1, u_2, u_3; h_{q_2}) ds d\mathbf{u},$$

where

$$(9.23) \quad \mathcal{Z}(s, u_1, u_2, u_3; h_{q_2}) = \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} \frac{\eta(h') Z(s, u_1, u_2, u_3; h' h_{q_2})}{(h')^{s/2+u_3}}.$$

**9.4. Properties of  $Z$  and  $\mathcal{Z}$ .** Define

$$\lambda_\pi^*(n) = \sum_{d|n} |\lambda_\pi(d)|.$$

Recalling  $\ell m = q_1 \frac{q_2}{h_{q_2}}$ , where  $(q_1, q_2) = 1$ , write

$$\ell = \ell_1 \ell_2, \quad m = m_1 m_2, \quad \text{where} \quad \ell_1 m_1 = q_1, \quad \ell_2 m_2 = \frac{q_2}{h_{q_2}},$$

and note  $(\ell_1 m_1, \ell_2 m_2) = 1$ . Recall also that  $(h', q_2) = 1$ ,  $q_1 | h'$ , that  $\eta$  has modulus  $q_2/h_{q_2}$ , and that  $q_2$  shares the same prime factors as  $q_2/h_{q_2}$ . In Lemma 9.2 below, we will also make use of the assumptions  $q_2^2 | d_2^3$  and  $d_2 | h_{q_2}$ .

Finally, we mention that  $\pi$  is an automorphic representation/newform of conductor  $m = m_1 m_2$  and central character  $\eta^2$ .

**Lemma 9.2.** *Let  $Z(s, u_1, u_2; h_{q_2} h')$  be defined by (9.20), initially with  $\text{Re}(s/2 + u_i)$  large,  $i = 1, 2$ . Then  $Z$  has a factorization  $Z = Z_{\text{good}} Z_{\text{bad}}$ , where*

$$Z_{\text{good}}(s, u_1, u_2; h_{q_2} h') = L(s/2 + u_1, \bar{\pi} \otimes \eta) L(s/2 + u_2, \bar{\pi} \otimes \eta),$$



and for  $\operatorname{Re}(s/2 + u_i) \geq \sigma/2 > 1/2$ ,  $i = 1, 2$  the series  $Z_{bad}$  is holomorphic, and we have

$$(9.24) \quad Z_{bad}(s, u_1, u_2; h_{q_2} h') \ll_{\sigma} (qN)^{\varepsilon} \left(\frac{h'}{q_1}, \ell_1\right)^{1/2-\theta} q_1 \ell_1 \ell_2^{1/2} \left(\frac{q_2}{h_{q_2}}\right)^{3/2} \lambda_{\pi}^* \left(\frac{h'}{q_1} \frac{h_{q_2}^3}{q_2^2}\right).$$

*Proof.* To begin, we recall the definition of  $\lambda_{\pi}^{(\delta)}$  from (6.23) and the estimate of Lemma 6.11(2), which together with Hecke multiplicativity imply

$$(9.25) \quad \lambda_{\pi}^{(\delta)}(n) \ll n^{\varepsilon} \sum_{d|(n, \delta)} d^{1/2} |\lambda_{\pi}(n/d)| \leq n^{\varepsilon} \sum_{d|(n, \delta)} d^{1/2} \lambda_{\pi}^*(n/d).$$

The function  $e \rightarrow \lambda_{\pi}^*(e)$  is approximately sub-multiplicative in the sense that  $\lambda_{\pi}^*(de) \ll_{\varepsilon} (de)^{\varepsilon} \lambda_{\pi}^*(d) \lambda_{\pi}^*(e)$ . Moreover,  $\lambda_{\pi}^*(d) \ll d^{\theta+\varepsilon} \ll d^{1/2}$ . These facts imply that the function  $d \rightarrow d^{1/2} \lambda_{\pi}^*(n/d)$  is essentially (i.e. up to a factor  $d^{\varepsilon}$ ) monotonically increasing in  $d$ . For example, we may deduce  $\lambda_{\pi}^{(\delta)}(n) \ll n^{\varepsilon} (n, \delta)^{1/2} \lambda_{\pi}^*\left(\frac{n}{(n, \delta)}\right)$ . This monotonicity property will be used repeatedly in the proof below.

The Dirichlet series  $Z = Z(s, u_1, u_2; h_{q_2} h')$  (see (9.20)) factors as follows. Write  $\delta$  uniquely as  $\delta = \delta_1 \delta_2$  with  $\delta_1 | \ell_1$  and  $\delta_2 | \ell_2$ , and let

$$h' = q_1 h_1 h'', \quad \text{where } h_1 | q_1^{\infty} \quad \text{and} \quad (h'', q) = 1.$$

Then  $Z = Z_0 Z_1 Z_2$  where

$$(9.26) \quad Z_0(s, u_1, u_2; h'') = \sum_{(n_1 n_2, q)=1} \frac{\eta(n_1 n_2)}{n_1^{\frac{s}{2}+u_1} n_2^{\frac{s}{2}+u_2}} \bar{\lambda}_{\pi}(h'') \bar{\lambda}_{\pi}(n_1 n_2),$$

$$Z_1(s, u_1, u_2; h_1) = \sum_{\delta_1 | \ell_1} \sum_{n_1, n_2 | q_1^{\infty}} \frac{S(0, n_1; q_1) S(0, n_2; q_1)}{n_1^{\frac{s}{2}+u_1} n_2^{\frac{s}{2}+u_2}} \eta(n_1 n_2) \bar{\lambda}_{\pi}^{(\delta_1)}(h_1) \bar{\lambda}_{\pi}^{(\delta_1)}(n_1 n_2),$$

and

$$(9.27) \quad Z_2(s, u_1, u_2; h_{q_2}) = \sum_{\delta_2 | \ell_2} \sum_{n_1, n_2 | q_2^{\infty}} \frac{\tau(\bar{\eta}) \tau(\bar{\eta}, n_1) \tau(\bar{\eta}, n_2)}{n_1^{\frac{s}{2}+u_1} n_2^{\frac{s}{2}+u_2}} \bar{\lambda}_{\pi}^{(\delta_2)} \left( n_1 n_2 \frac{h_{q_2}^2}{q_2} \right).$$

Let us begin with  $Z_0$ , for which it is not hard to see that

$$(9.28) \quad Z_0 = \frac{\bar{\lambda}_{\pi}(h'') L^{(q)}(s/2 + u_1, \bar{\pi} \otimes \eta) L^{(q)}(s/2 + u_2, \bar{\pi} \otimes \eta)}{\zeta^{(q)}(s + u_1 + u_2)},$$

where  $L^{(q)}(s, \pi)$  denotes the  $L$ -function  $L(s, \pi)$  with the Euler factors at the primes dividing  $q$  removed.

For  $Z_1$ , we claim

$$(9.29) \quad Z_1 \ll (h_1 q)^{\varepsilon} (h_1, \ell_1)^{1/2-\theta} q_1 \ell_1 \lambda_{\pi}^*(h_1),$$

as we now proceed to show. Let  $\sigma$  be such that  $\operatorname{Re}(s/2 + u_i) \geq \sigma/2 > 1/2$ . By the first estimate of (9.25),

$$(9.30) \quad Z_1 \ll h_1^{\varepsilon} \sum_{\delta_1 | \ell_1} \sum_{n_1, n_2 | q_1^{\infty}} \frac{(n_1, q_1)(n_2, q_1)}{(n_1 n_2)^{\sigma/2-\varepsilon}} \sum_{\substack{d|(h_1, \delta_1) \\ e|(n_1 n_2, \delta_1)}} (de)^{1/2} \left| \lambda_{\pi} \left( \frac{h_1}{d} \right) \right| \cdot \left| \lambda_{\pi} \left( \frac{n_1 n_2}{e} \right) \right|.$$

Since  $Z_1$  as well as its claimed upper bound (9.29) are multiplicative, it suffices to show (9.29) for  $q_1$  a prime power.

In the case that  $m_1 \neq 1$  we may use  $|\lambda_\pi(p)| \leq 1$  for  $p|m_1$  (see (6.1)) leading quickly to (9.29). In the case  $\ell_1 = q_1$ , we have by positivity the bound  $d^{1/2}|\lambda_\pi(h_1/d)| \leq (h_1, \ell_1)^{1/2}\lambda_\pi^*(h_1)$ , and the bound  $|\lambda_\pi(n_1 n_2/e)| \ll (n_1 n_2/e)^{\theta+\varepsilon}$  towards the Ramanujan conjecture. Currently any  $\theta > 7/64$  is admissible. Since  $q_1 = \ell_1$ , observe that the summand in (9.30) is maximized when  $n_1 = n_2 = e = \delta_1 = \ell_1$ . Bounding all terms by the  $e = \delta_1 = \ell_1$  term, the two preceding bounds show  $Z_1 \ll (h_1 q)^\varepsilon (h_1, \ell_1)^{1/2} \ell_1^{3/2+\theta} \lambda_\pi^*(h_1)$ ; we then derive (9.29) using  $(h_1, \ell_1)^{1/2} \ell_1^{3/2+\theta} \leq (h_1, \ell_1)^{1/2-\theta} \ell_1^2$ , since  $\frac{3}{2} + 2\theta \leq 2$ .

Next we estimate  $Z_2$ . We claim

$$(9.31) \quad Z_2 \ll (qh_{q_2})^\varepsilon \left(\frac{q_2}{h_{q_2}}\right)^{3/2} \ell_2^{1/2} \lambda_\pi^*(h_{q_2}^3/q_2^2).$$

The Dirichlet series  $Z_2$  can be factored into prime powers, provided that we correspondingly factor  $h$  and  $\eta$ ; the Gauss sums are multiplicative, up to a root of unity. Since the claimed bound on  $Z_2$  is multiplicative, and so is  $|Z_2|$ , it suffices to check it when  $q_2$  is a prime power.

We will use estimates for the Gauss sums from Section 7.1. Recall in particular from Corollary 7.2 that the product of Gauss sums  $\tau(\bar{\eta})\tau(\bar{\eta}, n_1)\tau(\bar{\eta}, n_2)$  vanishes except in the two cases

- (1)  $\eta$  is primitive modulo  $q_2/h_{q_2}$  and  $(n_1 n_2, q_2) = 1$
- (2)  $\eta$  is trivial, and  $q_2/h_{q_2} = p$ , prime.

Case 1) Suppose  $\eta$  is primitive modulo  $q_2/h_{q_2}$ . Then only the terms  $n_1 = n_2 = 1$  contribute to  $Z_2$ , and we have by the bound (9.25) for  $\lambda_\pi^{(\delta)}(n)$

$$(9.32) \quad Z_2 \ll \mathfrak{c}(\eta)^{3/2} q_2^\varepsilon \sum_{\delta_2|\ell_2} \sum_{d|(\delta_2, \frac{h_{q_2}^2}{q_2})} d^{1/2} \lambda_\pi^*\left(\frac{h_{q_2}^2/q_2}{d}\right).$$

Note that  $\ell_2 | \frac{q_2}{h_{q_2}}$  and  $\frac{q_2}{h_{q_2}} | \frac{h_{q_2}^2}{q_2}$ , since we recall  $q_2^2 | d_2^3$  and  $d_2 | h_{q_2}$ . Therefore, the largest value of  $d$  appearing in the bound (9.32) above is  $d = \ell_2$ . Since the summand in (9.32) is essentially monotonic, we have

$$Z_2 \ll q_2^\varepsilon \mathfrak{c}(\eta)^{3/2} \ell_2^{1/2} \lambda_\pi^*\left(\frac{h_{q_2}^2/q_2}{\ell_2}\right).$$

If  $m_2 \neq 1$ , then  $|\lambda_\pi(p)| \leq 1$  at primes dividing  $q_2$  (see (6.1)) and we easily obtain (9.31). If  $m_2 = 1$ , then  $\ell_2 = q_2/h_{q_2}$ , and the bound simplifies as

$$Z_2 \ll q_2^\varepsilon \mathfrak{c}(\eta)^{3/2} \ell_2^{1/2} \lambda_\pi^*(h_{q_2}^3/q_2^2).$$

Substituting  $\mathfrak{c}(\eta) = q_2/h_{q_2}$  gives the desired bound.

Case 2) Suppose  $\eta$  is the trivial character and  $q_2/h_{q_2} = p$  is prime. We obtain by (9.25)

$$(9.33) \quad Z_2 \ll q_2^\varepsilon \sum_{\delta_2|\ell_2} \sum_{n_1, n_2|p^\infty} \frac{(n_1, p)(n_2, p)}{(n_1 n_2)^{\sigma/2-\varepsilon}} \sum_{d|(\delta_2, n_1 n_2 \frac{h_{q_2}^2}{q_2})} d^{1/2} \lambda_\pi^*\left(\frac{n_1 n_2 h_{q_2}^2/q_2}{d}\right).$$

The largest value of  $d$  appearing in the above sum is at most  $\ell_2$  regardless of  $n_1, n_2$ , and the summand is monotonic increasing in  $d$ . Similarly, the summand in (9.33) is monotonically decreasing as a function of  $n_1, n_2$  as soon as  $n_1, n_2 \geq p$  by current progress towards

Ramanujan. Hence

$$Z_2 \ll q_2^\varepsilon p \ell_2^{1/2} \lambda_\pi^* \left( \frac{p^2 h_{q_2}^2 / q_2}{\ell_2} \right) = q_2^\varepsilon p \ell_2^{1/2} \lambda_\pi^* \left( \frac{q_2}{\ell_2} \right).$$

We can further simplify this bound as follows. If  $m_2 \neq 1$ , then  $|\lambda_\pi(p)| \leq 1$  for  $p|m_2$ . If  $m_2 = 1$ , then  $\ell_2 = q_2/h_{q_2} = p$ . In either case, we obtain

$$Z_2 \ll q_2^\varepsilon p \ell_2^{1/2} \lambda_\pi^*(h_{q_2}).$$

This bound can in turn be absorbed by (9.31) using current progress towards Ramanujan, since  $h_{q_2} = \frac{h_{q_2}^3}{q_2^2} \frac{q_2^2}{h_{q_2}^2}$ , so that  $\lambda_\pi^*(h_{q_2}) \leq \lambda_\pi^*(h_{q_2}^3/q_2^2)(q_2/h_{q_2})^{2\theta}$ , and  $1 + 2\theta \leq 3/2$ .

Putting together the previous estimates finishes the proof.  $\square$

**Lemma 9.3.** *Suppose that  $q = q_1 q_2$  with  $(q_1, q_2) = 1$ , and suppose that  $h_{q_2}$  is an integer as in (9.21). Then*

$$(9.34) \quad \mathcal{Z}(s, u_1, u_2, u_3; h_{q_2}) = L(s/2 + u_1, \bar{\pi} \otimes \eta) L(s/2 + u_2, \bar{\pi} \otimes \eta) L(s/2 + u_3, \bar{\pi} \otimes \eta) \mathcal{Z}_{bad},$$

where for  $\operatorname{Re}(s/2 + u_i) \geq \sigma/2 > 1/2$ ,  $i = 1, 2, 3$  the series  $\mathcal{Z}_{bad}$  is holomorphic, and we have

$$(9.35) \quad \mathcal{Z}_{bad} \ll_\sigma (qN)^\varepsilon q_1^{-\sigma/2} q_1 \ell_1 \ell_2^{1/2} \left( \frac{q_2}{h_{q_2}} \right)^{3/2} \lambda_\pi^* \left( \frac{h_{q_2}^3}{q_2^2} \right).$$

Remark. Lemma 9.3 implies that  $\mathcal{Z}$  has analytic continuation to the region of  $\mathbb{C}^4$  with  $\operatorname{Re}(s/2 + u_i) > 1/2$ ,  $i = 1, 2, 3$ .

*Proof.* As in the proof of Lemma 9.2, we have a factorization of the left hand side of (9.34) as  $\mathcal{Z}_0 \mathcal{Z}_1 \mathcal{Z}_2$ , analogously to (9.26)–(9.27). For instance,

$$\mathcal{Z}_1 = \sum_{h_1 | q_1^\infty} \frac{\eta(q_1 h_1)}{(q_1 h_1)^{s/2 + u_3}} Z_1(s, u_1, u_2; h_1).$$

Then using (9.29), we have

$$\mathcal{Z}_1 \ll q_1^{\varepsilon - \sigma/2} q_1 \ell_1 \sum_{h_1 | q_1^\infty} \frac{\lambda_\pi^*(h_1)}{h_1^{\sigma/2}} (h_1, \ell_1)^{1/2 - \theta} \ll q_1^{\varepsilon - \sigma/2} q_1 \ell_1.$$

The case of  $\mathcal{Z}_2$  is easy now, because  $\mathcal{Z}_2$  is identical to (9.27), since  $(h', q_2) = 1$ . Therefore, the bound (9.31) holds for  $\mathcal{Z}_2$ . Finally, it is easy to see from (9.28) that

$$\mathcal{Z}_0 = \sum_{(h'', q_2)=1} \frac{\eta(h'')}{(h'')^{s/2 + u_3}} Z_0(s, u_1, u_2, u_3; h'') = \frac{\prod_{j=1}^3 L^{(q)}(s/2 + u_j, \bar{\pi} \otimes \eta)}{\zeta^{(q)}(s + u_1 + u_2)}. \quad \square$$

**9.5. Properties of  $Z$ , Eisenstein case.** Suppose that  $\eta$  is a Dirichlet character modulo  $q_2/h_{q_2}$ , which we may also identify with a character of  $\widehat{\mathbb{Z}}^\times$ . For a Hecke character  $\mu$ , we write  $\mu \otimes \eta$  for the twist of  $\mu$  by the finite order Hecke character corresponding to  $\eta$ . Let  $Z$  be defined by (9.20), but where  $\pi$  is a global principal series/newform Eisenstein series of conductor  $m$ , central character  $\eta^2$ , and spectral parameter  $it$ . Similarly define  $\mathcal{Z}$  for  $\pi$  a global principal series/newform Eisenstein series.

**Lemma 9.4.** *The following properties hold for  $Z$  and  $\mathcal{Z}$  in the Eisenstein case:*

- (1) *Lemmas 9.2 and 9.3 carry over verbatim.*

(2) The series  $\mathcal{Z}$  has meromorphic continuation to the region of  $\mathbb{C}^4$  with  $\operatorname{Re}(s/2 + u_i) > 1/2$ ,  $i = 1, 2, 3$ . Its only possible poles are at  $s/2 + u_i \pm it = 1$ , which occur if and only if  $\pi \simeq \pi(\mu_1, \mu_2)$  with  $\mu_i$  unitary Hecke characters of conductors  $r_i$ ,  $i = 1, 2$  such that  $\eta = \mu_1|_{\widehat{\mathbb{Z}}^\times} = \mu_2|_{\widehat{\mathbb{Z}}^\times}$ .

In particular, only if  $r_1 = r_2$  and  $r_1 r_2 = m$ , so  $m_1 = 1$ , and  $m$  is a square.

(3) Factoring  $\mathcal{Z} = \mathcal{Z}_{\text{good}} \mathcal{Z}_{\text{bad}}$  as in (9.34), for  $\operatorname{Re}(s/2 + u_i) \geq 1 + \varepsilon$ ,  $i = 1, 2, 3$ , we have

$$(9.36) \quad \mathcal{Z}_{\text{bad}} \ll (qN)^\varepsilon q_1^{-1} \ell_1 \ell_2^{1/2} \mathbf{c}(\eta)^{3/2}.$$

*Proof.* Part (1) follows from inspection of the proofs of Lemmas 9.2 and 9.3. This also gives the meromorphic continuation of  $\mathcal{Z}$  with possible poles only at the poles of  $L(s/2 + u_i, \bar{\pi} \otimes \eta)$  for  $i = 1, 2, 3$ .

Write  $\pi = \pi(\mu_1, \mu_2)$  where  $\mu_1, \mu_2$  have conductors  $r_1, r_2$ , respectively. Therefore,  $\mu_1 \mu_2 = \eta^2$ ,  $r_1 r_2 = m$ , and

$$L(s, \bar{\pi} \otimes \eta) = L(s, \bar{\mu}_1 \otimes \eta) L(s, \bar{\mu}_2 \otimes \eta).$$

This  $L$ -function has a pole if and only if the conductor  $\mathbf{c}(\bar{\mu}_1 \otimes \eta) = 1$ , equivalently if and only if  $\mathbf{c}(\bar{\mu}_2 \otimes \eta) = 1$ . In this case, since  $\bar{\mu}_1 \otimes \eta$  is unramified at all finite places and  $\eta$  is finite order, we have  $\eta = \mu_1|_{\widehat{\mathbb{Z}}^\times} = \mu_2|_{\widehat{\mathbb{Z}}^\times}$ . In particular, the possible locations of the poles are as stated in the lemma. Moreover, since  $\eta$  has conductor dividing  $q_2/h_{q_2}$  (which is coprime to  $q_1$ ), this implies  $m_1 = 1$ . Additionally,  $r_1 = r_2$ , and so  $m = r_1^2$ .

Finally, to show (9.36), we simply revisit the estimates of  $\mathcal{Z}_1$  and  $\mathcal{Z}_2$  from the proofs of Lemmas 9.2 and 9.3, but now the estimates occur slightly to the right of the 1-line instead of slightly to the right of the 1/2-line, and we may additionally use the Ramanujan bound for simplicity. Indeed, inspection of (9.30) shows  $Z_1 \ll (h_1 q)^\varepsilon \ell_1$ , and hence  $\mathcal{Z}_1 \ll (h_1 q)^\varepsilon q_1^{-1} \ell_1$ . Similarly, (9.32) and (9.33) lead to  $\mathcal{Z}_2 = Z_2 \ll q_2^\varepsilon \mathbf{c}(\eta)^{3/2} \ell_2^{1/2}$ .  $\square$

## 9.6. Using the spectral bounds.

**Proposition 9.5.** *With notation as in this section, we have*

$$(9.37) \quad \sum_{h \equiv 0 \pmod{d}} S_{\text{Maass}}(\chi, h) \ll N(qN)^\varepsilon.$$

Observe that Proposition 9.5 is consistent with Theorem 1.5.

*Proof.* We take the expression (9.22), move the contours of integration so  $\operatorname{Re}(s) = 1 + \varepsilon$ ,  $\operatorname{Re}(u_j) = \varepsilon$ ,  $j = 1, 2, 3$ , and apply the triangle inequality. We use (9.17) and Lemma 9.3 to bound  $H_2$  and  $\mathcal{Z}$ , respectively. Altogether, we obtain

$$(9.38) \quad \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h') \ll \frac{h_{q_2}^2 q_1^{-1/2} (qN)^\varepsilon}{q^2 \varphi(q_2/h_{q_2})} \sum_{\eta \pmod{q_2/h_{q_2}}} \frac{h_{q_2}}{q} \left(\frac{q}{h_{q_2}^2}\right)^{1/2} \frac{Nq}{h_{q_2}} \sum_{t_j \ll (qN)^\varepsilon} \sum_{\ell m = \frac{q}{h_{q_2}}} \\ q_1 \ell_1 \ell_2^{1/2} \left(\frac{q_2}{h_{q_2}}\right)^{3/2} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \lambda_\pi^* \left(\frac{h_{q_2}^3}{q_2^2}\right) \int \prod_{i=1}^3 |L(s/2 + u_i, \bar{\pi} \otimes \eta)| ds d\mathbf{u},$$

plus a small error term. The limits on the integral sign are not displayed; to be definite, the integrals (over  $s, u_1, u_2, u_3$ ) have real parts as fixed above. Since the function  $H_2$  has rapid

decay along vertical lines, the integrals may be truncated at  $(qN)^\varepsilon$  with a very small error term.

Before applying any more advanced tools to bound this expression, we first clean it up with trivial simplifications, giving

$$(9.39) \quad \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h') \ll \frac{h_{q_2}^{1/2} N(qN)^\varepsilon}{q} \sum_{\eta \pmod{q_2/h_{q_2}}} \sum_{t_j \ll (qN)^\varepsilon} \\ \sum_{\ell m = \frac{q}{h_{q_2}}} \ell_1 \ell_2^{1/2} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \lambda_\pi^* \left( \frac{h_{q_2}^3}{q_2^2} \right) \int \prod_{i=1}^3 |L(s/2 + u_i, \bar{\pi} \otimes \eta)| ds du,$$

plus a small error term. Inspired by the method of [BM], we apply Hölder's inequality with exponents  $(4, 4, 4, 4)$ . Note that  $(\lambda_\pi^*(m))^2 \ll m^\varepsilon \lambda_\pi^*(m^2)$ , using the Hecke relations. This gives

$$\sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h') \ll \frac{h_{q_2}^{1/2} N(qN)^\varepsilon}{q} A_1^{1/4} A_2^{1/4} A_3^{1/4} A_4^{1/4},$$

where

$$A_1 = \sum_{t_j \ll (qN)^\varepsilon} \sum_{\ell m = \frac{q}{h_{q_2}}} \ell_1 \ell_2^2 \sum_{\eta \pmod{q_2/h_{q_2}}} \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \left| \lambda_\pi^* \left( \frac{h_{q_2}^6}{q_2^4} \right) \right|^2,$$

and for  $i = 2, 3, 4$ ,

$$A_i = \sum_{\eta \pmod{q_2/h_{q_2}}} \sum_{t_j \ll (qN)^\varepsilon} \sum_{\ell m = \frac{q}{h_{q_2}}} \ell_1 \sum_{\pi \in \mathcal{H}_{it_j}(m, \eta^2)} \int |L(s/2 + u_i, \bar{\pi} \otimes \eta)|^4 ds du_i.$$

Note that we arranged  $\ell_1$  evenly in each  $A_i$ , but took  $(\ell_2^{1/2})^4 = \ell_2^2$  in  $A_1$ .

Now we turn to the estimates for each  $A_i$ , starting with  $A_1$ . Our aim is to apply Lemma 7.5, but  $A_1$  is not quite in the correct form. This can be easily remedied by arranging

$$\{\eta \pmod{q_2/h_{q_2}}\} = \bigcup_{\substack{\psi \pmod{q_2/h_{q_2}} \\ \text{even}}} \{\eta : \eta^2 = \psi\}.$$

Note that the interior set on the right hand side is of cardinality  $\ll q_2^\varepsilon$ , since it is a coset of the subgroup of characters modulo  $q_2/h_{q_2}$  of order dividing 2. With this observation, Lemma 7.5 gives

$$A_1 \ll (qN)^\varepsilon \sum_{\ell_1 m_1 = q_1} \ell_1 \sum_{\ell_2 m_2 = \frac{q_2}{h_{q_2}}} \ell_2^2 \left( m_1 m_2^2 + \left( \frac{h_{q_2}^3}{q_2^2} \right) m_2^{1/2} \right),$$

Recalling that  $(q_2/h_{q_2})^2 \leq h_{q_2}$ , we deduce

$$A_1 \ll (qN)^\varepsilon q_1 h_{q_2}.$$

For  $A_i$ ,  $i = 2, 3, 4$ , by Theorem 7.6 we have

$$A_i \ll (qN)^\varepsilon \frac{q_1 q_2^2}{h_{q_2}^2}.$$

Therefore,

$$(9.40) \quad \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Maass}}(\chi, h_{q_2} h') \ll \frac{h_{q_2}^{1/2} N (qN)^\varepsilon}{q} q_1^{1/4} h_{q_2}^{1/4} \left( \frac{q_1 q_2^2}{h_{q_2}^2} \right)^{3/4} = (qN)^\varepsilon N \frac{q_2^{1/2}}{h_{q_2}^{3/4}}.$$

Finally, we insert (9.40) into (9.21), giving

$$\sum_{\substack{h \ll H \\ h \equiv 0 \pmod{d}}} S_{\text{Maass}}(\chi, h) \ll (qN)^\varepsilon N \sum_{\substack{q_1 q_2 = q \\ (q_1, q_2) = 1}} \sum_{\substack{h_{q_2} | q_2 \\ v_p(h_{q_2}) < v_p(q_2) \\ d_2 | h_{q_2}}} \frac{q_2^{1/2}}{h_{q_2}^{3/4}}.$$

The summand is largest when  $h_{q_2} = d_2$ , and since  $q_2^2 \mid d_2^3$ , we have  $d_2^{3/4} \geq q_2^{1/2}$ , which completes the proof of Proposition 9.5.  $\square$

**Proposition 9.6.** *The bound stated in Proposition 9.5 holds for  $S_{\text{hol}}$ .*

*Proof.* The proof is nearly identical, changing the obvious things that need to be changed.  $\square$

**9.7. The Eisenstein series contribution.** The contribution of the Eisenstein series  $S_{\text{Eis}}$  is largely similar to the Maass form case, with one important difference. In the Maass case, we related  $K_{\text{Maass}}$  to a Mellin integral involving twisted  $L$ -functions, which we shifted slightly to the right of the critical line. In the Eisenstein case, we will perform the same steps, but there will be a polar contribution arising since the twisted  $L$ -functions are now products of Dirichlet  $L$ -functions, which may have a pole when the character is trivial. We denote by  $S_{\text{Pole}}$  the contribution from this pole.

**Proposition 9.7.** *The same bound stated in Proposition 9.5 holds for  $S_{\text{Eis}} - S_{\text{Pole}}$ .*

*Proof.* This is clear from Lemma 9.4.  $\square$

**Proposition 9.8.** *We have*

$$S_{\text{Pole}}(\chi) \ll d^{-1/8} \frac{NH}{q} (Nq)^\varepsilon.$$

This bound is more than satisfactory for Theorem 1.5.

*Proof.* The term  $S_{\text{Pole}}(\chi)$  only arises when  $\mathcal{Z}$  has a pole. By Lemma 9.4, such poles arise at  $\frac{s}{2} + u_i \pm it = 1$  precisely when  $\pi \simeq \pi(\mu_1, \mu_2)$  with  $\eta = \mu_1|_{\widehat{\mathbb{Z}}^\times} = \mu_2|_{\widehat{\mathbb{Z}}^\times}$ . In particular,  $\mathbf{c}(\pi) = \mathbf{c}(\eta)^2$ , and  $\ell_2 \mathbf{c}(\eta)^2 = \frac{q_2}{h_{q_2}}$ . The contribution of the polar terms may be estimated by setting the integrals in the Eisenstein analogue of (9.22) to  $\text{Re}(s) = 1 + \varepsilon$ , and  $\text{Re}(u_i) = 1/2 + \varepsilon$  for  $i = 1, 2, 3$ , and using (9.36) to estimate  $\mathcal{Z}_{\text{bad}}$  in lieu of (9.35). Writing  $\mathbf{c}(\eta) = v$ , and using (9.17) we obtain

$$\begin{aligned} \sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Pole}}(\chi, h_{q_2} h') &\ll \frac{h_{q_2}^2}{q^2 \frac{q_2}{h_{q_2}}} \sum_{\ell_2 v^2 = \frac{q_2}{h_{q_2}}} \frac{\ell_1}{q_1} \ell_2^{1/2} v^{3/2} \\ &\times \sum_{\eta \pmod{v}}^* \frac{(q/h_{q_2}^2)^{1/2}}{q/h_{q_2}} \left( \frac{Nq_2}{h_{q_2}} \right) \frac{(M_1 M_2 H)^{1/2}}{h_{q_2}^{3/2}} (Nq)^\varepsilon. \end{aligned}$$

Recalling that  $M_1 M_2 = q^2$  (see (9.18)) and simplifying gives

$$\sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Pole}}(\chi, h_{q_2} h') \ll \frac{NH^{1/2} h_{q_2}^{1/2} (Nq)^\varepsilon}{q^{3/2}} \sum_{\ell_2 v^2 = \frac{q_2}{h_{q_2}}} \ell_2^{1/2} v^{5/2}.$$

Note  $\ell_2^{1/2} v^{5/2} \leq (q_2/h_{q_2})^{5/4}$ . Therefore,

$$\sum_{\substack{h' \equiv 0 \pmod{q_1} \\ (h', q_2) = 1}} S_{\text{Pole}}(\chi, h_{q_2} h') \ll \frac{NH^{1/2} (Nq)^\varepsilon}{q_1^{3/2} q_2^{1/4} h_{q_2}^{3/4}}.$$

Including the outer summations appearing in (9.21), we then deduce

$$S_{\text{Pole}}(\chi) \ll \sum_{\substack{q_1 q_2 = q \\ (q_1, q_2) = 1}} \frac{NH^{1/2} (Nq)^\varepsilon}{q_1^{3/2} q_2^{1/4} d_2^{3/4}} = \frac{NH(Nq)^\varepsilon}{q} \frac{q_2^{3/4}}{q_1^{1/2} H^{1/2} d_2^{3/4}}.$$

Using  $H \geq q_1 d_2$  (since  $h \ll H$  and  $h \equiv 0 \pmod{q_1 d_2}$ ) and  $q_2 \leq d_2^{3/2}$  shows the claimed bound.  $\square$

Applying Propositions 9.5, 9.6, 9.7, 9.8 and (8.8) to the sum over  $h \equiv 0 \pmod{d}$  of (9.4), we conclude the proof of Theorem 1.5.

## REFERENCES

- [AK] N. Andersen and E. M. Kiral. Level reciprocity in the twisted second moment of Rankin-Selberg  $L$ -functions. *Mathematika*, 64(3):770–784, 2018.
- [AtLe] A. O. L. Atkin and J. Lehner. Hecke operators on  $\Gamma_0(m)$ . *Math. Ann.*, 185:134–160, 1970.
- [AtLi] A. O. L. Atkin and W.-Ch. W. Li. Twists of newforms and pseudo-eigenvalues of  $W$ -operators. *Invent. Math.*, 48(3): 221–243, 1978.
- [BH] Colin J. Bushnell and Guy Henniart. Higher ramification and the local Langlands correspondence. *Ann. of Math. (2)*, 185(3):919–955, 2017.
- [BHKM] V. Blomer, P. Humphries, R. Khan, and M. Milinovich. Motohashi’s fourth moment identity for non-archimedean test functions and applications. *Compos. Math.*, 156(5):1004–1038, 2020.
- [BK] V. Blomer and R. Khan. Twisted moments of  $L$ -functions and spectral reciprocity. *Duke Math J.*, 168(6):1109–1177, 2019.
- [BM] V. Blomer and Dj. Milićević. The second moment of twisted modular  $L$ -functions. *Geom. Funct. Anal.*, 25(2):453–516, 2015.
- [BLS] A. R. Booker, M. Lee, and A. Strömbergsson. Twist-minimal trace formulas and the Selberg eigenvalue conjecture. *J. Lond. Math. Soc. (2)*, no. 3, 1067–1134.
- [B] D. Bump. *Automorphic forms and representations*, volume 55 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.
- [C] W. Casselman. On some results of Atkin and Lehner. *Math. Ann.*, 201:301–314, 1973.
- [CI] J. B. Conrey and H. Iwaniec. The cubic moment of central values of automorphic  $L$ -functions. *Ann. of Math. (2)*, 151(3):1175–1216, 2000.
- [CS] A. Corbett and A. Saha. On the order of vanishing of newforms at cusps. *Math. Res. Lett.*, 25(6):1771–1804, 2018.
- [Da] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000.
- [De1] P. Deligne. Formes modulaires et représentations de  $GL(2)$ . In *Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*, pages 55–105. Lecture Notes in Math., Vol. 349. Springer, Berlin, 1973.

- [De2] P. Deligne. *Cohomologie étale*, volume 569 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1977. Séminaire de géométrie algébrique du Bois-Marie SGA 4 $\frac{1}{2}$ .
- [DI] J.-M. Deshouillers and H. Iwaniec. Kloosterman sums and Fourier coefficients of cusp forms. *Invent. Math.*, 70(2):219–288, 1982/83.
- [DFI] W. Duke, J. B. Friedlander, and H. Iwaniec. A quadratic divisor problem. *Invent. Math.*, 115(2):209–217, 1994.
- [Fr] D. Frolenkov. The cubic moment of automorphic  $L$ -functions in the weight aspect. *J. Number Theory*, 207:247–281, 2020.
- [GR] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007.
- [G] J. Guo. On the positivity of the central critical values of automorphic  $L$ -functions for  $GL(2)$ . *Duke Math. J.*, 83(1):157–190, 1996.
- [H-B1] D. R. Heath-Brown. Hybrid bounds for Dirichlet  $L$ -functions. *Invent. Math.*, 47(2):149–170, 1978.
- [H-B2] D. R. Heath-Brown. The fourth power moment of the Riemann zeta function. *Proc. London Math. Soc. (3)*, 38(3):385–422, 1979.
- [HL] J. Hoffstein and P. Lockhart. Coefficients of Maass forms and the Siegel zero. *Ann. of Math. (2)*, 140(1):161–181, 1994. With an appendix by Dorian Goldfeld, Hoffstein and Daniel Lieman.
- [Iv] A. Ivić. On sums of Hecke series in short intervals. *J. Théor. Nombres Bordeaux*, 13(2):453–468, 2001.
- [IK] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
- [ILS] H. Iwaniec, W. Luo, and P. Sarnak. Low lying zeros of families of  $L$ -functions. *Inst. Hautes Études Sci. Publ. Math.*, (91):55–131 (2001), 2000.
- [Iw1] H. Iwaniec. Fourier coefficients of cusp forms and the Riemann zeta-function. In *Seminar on Number Theory, 1979–1980 (French)*, pages Exp. No. 18, 36. Univ. Bordeaux I, Talence, 1980.
- [Iw2] H. Iwaniec. Small eigenvalues of Laplacian for  $\Gamma_0(N)$ . *Acta Arith.*, 56(1):65–82, 1990.
- [Iw3] H. Iwaniec. *Topics in classical automorphic forms*, volume 17 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.
- [Iw4] H. Iwaniec. *Spectral methods of automorphic forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, second edition, 2002.
- [JL] H. Jacquet and R. P. Langlands. *Automorphic forms on  $GL(2)$* , volume 114 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York, 1970.
- [JM1] M. Jutila and Y. Motohashi. Mean value estimates for exponential sums and  $L$ -functions: a spectral-theoretic approach. *J. Reine Angew. Math.*, 459:61–87, 1995.
- [JM2] M. Jutila and Y. Motohashi. Uniform bound for Hecke  $L$ -functions. *Acta Math.*, 195:61–115, 2005.
- [KMN] R. Khan, Dj. Milićević, and H. T. Ngo. Non-vanishing of Dirichlet  $L$ -functions in Galois orbits. *Int. Math. Res. Not. IMRN*, (22):6955–6978, 2016.
- [KY1] E. M. Kiral and M. P. Young. The fifth moment of modular  $L$ -functions. *J. Eur. Math. Soc. (JEMS)* 23 (2021), no. 1, 237–314.
- [KY2] E. M. Kiral and M. P. Young. Kloosterman sums and Fourier coefficients of Eisenstein series. *Ramanujan J.*, 49(2):391–409, 2019.
- [KL1] A. Knightly and C. Li. *Traces of Hecke operators*, volume 133 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.
- [KL2] A. Knightly and C. Li. Kuznetsov’s trace formula and the Hecke eigenvalues of Maass forms. *Mem. Amer. Math. Soc.*, 224(1055):vi+132, 2013.
- [KMS] E. Kowalski, Ph. Michel, and W. Sawin. Bilinear forms with Kloosterman sums and applications. *Ann. of Math. (2)*, 186(2):413–500, 2017.
- [La] E. Lapid. On an inequality of Bushnell-Henniart for Rankin-Selberg conductors. *Israel J. Math.*, 2021.
- [Li1] W.-Ch. W. Li. Newforms and functional equations. *Math. Ann.*, 212:285–315, 1975.
- [Li2] W.-Ch. W. Li.  $L$ -series of Rankin type and their functional equations. *Math. Ann.*, 244(2):135–166, 1979.
- [MV] Ph. Michel and A. Venkatesh. The subconvexity problem for  $GL_2$ . *Publ. Math. Inst. Hautes Études Sci.*, (111):171–271, 2010.



- [Mi] Dj. Milićević. Sub-Weyl subconvexity for Dirichlet  $L$ -functions to prime power moduli. *Compos. Math.*, 152(4):825–875, 2016.
- [MW] Dj. Milićević and D. White. Twelfth moment of Dirichlet  $L$ -functions to prime power moduli. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 22(4):1879–1898, 2021.
- [Mo] Y. Motohashi. *Spectral theory of the Riemann zeta-function*, volume 127 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.
- [NPS] P. D. Nelson, A. Pitale, and A. Saha. Bounds for Rankin-Selberg integrals and quantum unique ergodicity for powerful levels. *J. Amer. Math. Soc.*, 27(1):147–191, 2014.
- [Nel] P. D. Nelson. Eisenstein series and the cubic moment for  $\mathrm{PGL}(2)$ . arXiv:1911.06310, 2019.
- [Neu] J. Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1999.
- [Nun] R. M. Nunes. The twelfth moment of Dirichlet  $L$ -functions with smooth moduli. *Int. Math. Res. Not. IMRN*, (12):9180–9202, 2021.
- [P1] I. Petrow. A twisted Motohashi formula and Weyl-subconvexity for  $L$ -functions of weight two cusp forms. *Math. Ann.*, 363(1-2): 175–216, 2015.
- [P2] I. Petrow. Bounds for traces of Hecke operators and applications to modular and elliptic curves over a finite field. *Algebra Number Theory*, 12(10):2471–2498, 2018.
- [PY1] I. Petrow and M. P. Young. A generalized cubic moment and the Petersson formula for newforms. *Math. Ann.*, 373(1-2):287–353, 2019.
- [PY2] I. Petrow and M. P. Young. The Weyl bound for Dirichlet  $L$ -functions of cube-free conductor. *Ann. of Math. (2)*, 192 (2020), no. 2, 437–486.
- [SST] P. Sarnak, S. Shin, and N. Templier. Families of  $L$ -functions and their symmetry. *Families of automorphic forms and the trace formula*, Simons Symp., Springer, [Cham], 2016.
- [Sch] R. Schmidt. Some remarks on local newforms for  $\mathrm{GL}(2)$ . *J. Ramanujan Math. Soc.*, 17(2):115–147, 2002.
- [SPY] R. Schulze-Pillot and A. Yenirce. Petersson products of bases of spaces of cusp forms and estimates for Fourier coefficients. *Int. J. Number Theory*, 14(8):2277–2290, 2018.
- [Se] J.-P. Serre. *A course in arithmetic*, volume 7 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Heidelberg, 1973.
- [Sm] R. A. Smith. On  $n$ -dimensional Kloosterman sums. *J. Number Theory*, 11:324–343, 1979.
- [So] K. Soundararajan. The fourth moment of Dirichlet  $L$ -functions. In *Analytic number theory*, volume 7 of *Clay Math. Proc.*, pages 239–246. Amer. Math. Soc., Providence, RI, 2007.
- [T] J. B. Tunnell. On the local Langlands conjecture for  $\mathrm{GL}(2)$ . *Invent. Math.*, 46(2):179–200, 1978.
- [Y1] M. P. Young. The fourth moment of Dirichlet  $L$ -functions. *Ann. of Math. (2)*, 173(1):1–50, 2011.
- [Y2] M. P. Young. Explicit calculations with Eisenstein series. *J. Number Theory*, 199:1–48, 2019.
- [Z] R. Zacharias. Periods and reciprocity I. *Int. Math. Res. Not. IMRN*, 2021, no. 3, 2191–2209.

*Email address:* `ian.petrow@math.ethz.ch`

ETH ZÜRICH, DEPARTMENT OF MATHEMATICS, RÄMISTRASSE 101, 8092 ZÜRICH, SWITZERLAND

*Email address:* `myoung@math.tamu.edu`

DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77843-3368, U.S.A.