

RESEARCH ARTICLE

# Bounds for twisted symmetric square $L$ -functions via half-integral weight periods

Paul D. Nelson

ETH Zürich, Department of Mathematics, Rämistrasse 101, CH-8092, Zürich, Switzerland; E-mail: [paul.nelson@math.ethz.ch](mailto:paul.nelson@math.ethz.ch).

Received: 23 January 2020; Revised: 22 May 2020; Accepted: 11 June 2020

2020 Mathematics Subject Classification: Primary 11F27; Secondary 11F37, 58J51

Keywords and phrases:  $L$ -functions; for each term replace; and give space and; modular forms, half-integral weight, quadratic twists

## Abstract

We establish the first moment bound

$$\sum_{\varphi} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) \ll_{\varepsilon} p^{5/4+\varepsilon}$$

for triple product  $L$ -functions, where  $\Psi$  is a fixed Hecke–Maass form on  $SL_2(\mathbb{Z})$  and  $\varphi$  runs over the Hecke–Maass newforms on  $\Gamma_0(p)$  of bounded eigenvalue. The proof is via the theta correspondence and analysis of periods of half-integral weight modular forms. This estimate is not expected to be optimal, but the exponent  $5/4$  is the strongest obtained to date for a moment problem of this shape. We show that the expected upper bound follows if one assumes the Ramanujan conjecture in both the integral and half-integral weight cases.

Under the triple product formula, our result may be understood as a strong level aspect form of quantum ergodicity: for a large prime  $p$ , all but very few Hecke–Maass newforms on  $\Gamma_0(p)\backslash\mathbb{H}$  of bounded eigenvalue have very uniformly distributed mass after pushforward to  $SL_2(\mathbb{Z})\backslash\mathbb{H}$ .

Our main result turns out to be closely related to estimates such as

$$\sum_{|n|<p} L(\Psi \otimes \chi_{np}, \frac{1}{2}) \ll p,$$

where the sum is over those  $n$  for which  $np$  is a fundamental discriminant and  $\chi_{np}$  denotes the corresponding quadratic character. Such estimates improve upon bounds of Duke–Iwaniec.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Overview	2
1.2	Context	2
1.3	Main result	3
1.4	Conditional sharp bounds	4
1.5	Application to moments of triple product $L$ -functions	4
1.6	Application to sparse moments of quadratic twists	5
1.7	Method	6

<b>2</b>	<b>Division of the proof</b>	<b>7</b>
2.1	Jacobi theta function . . . . .	7
2.2	Theta multiplier . . . . .	7
2.3	Petersson inner product . . . . .	7
2.4	The half-integral weight lift and its Fourier expansion . . . . .	7
2.5	Application of an incomplete Hecke operator . . . . .	8
2.6	Properties of the varying forms . . . . .	8
2.7	An explicit seesaw identity . . . . .	9
2.8	Reduction to period bounds . . . . .	9
2.9	The basic inequality . . . . .	9
2.10	Completion of the proof . . . . .	10
<b>3</b>	<b>Preliminaries</b>	<b>10</b>
3.1	Generalities . . . . .	10
3.2	Theta kernels . . . . .	11
3.3	Ternary theta lifts . . . . .	12
3.4	Cusps of $\Gamma_0(4/p)$ . . . . .	13
3.5	Explicit Shimizu lifts . . . . .	14
<b>4</b>	<b>The <math>L^2</math>-norm of <math>h^\sharp</math></b>	<b>14</b>
<b>5</b>	<b>A triple product identity</b>	<b>16</b>
5.1	Linear algebra lemma . . . . .	16
5.2	Pushforward of the theta kernel . . . . .	17
5.3	Period identities . . . . .	17
<b>6</b>	<b>The basic inequality</b>	<b>19</b>
<b>7</b>	<b>A converse estimate</b>	<b>19</b>

## 1. Introduction

### 1.1. Overview

The quantum ergodicity theorem [35, 4, 44] says that on a compact Riemannian manifold with ergodic geodesic flow, almost all eigenfunctions have equidistributed mass in the large eigenvalue limit. When the manifold is arithmetic, additional tools become available by which one can prove quantitative strengthenings of this conclusion, to the effect that all but *very few* eigenfunctions (satisfying additional symmetries) have *very* equidistributed mass (see Section 1.2 below, or [21, 22, 17]). A standard way to quantify such strengthenings is through upper bounds for  $L^2$ -mass variance over families. Le Masson and Sahlsten [19] recently introduced a level aspect variant of the quantum ergodicity theorem concerning almost all eigenfunctions in a fixed spectral window on a sequence of hyperbolic surfaces Benjamini–Schramm-converging to the hyperbolic plane. The main results of this article may be understood as quantitative strengthenings of that result, for specific classes of eigenfunctions and observables, in the arithmetic congruence case.

Well-developed techniques for analyzing averages of triple product  $L$ -values and/or shifted convolution sums apply in our setting, giving nontrivial estimates in the intended direction. We do not apply such techniques here. We instead introduce techniques involving the theta correspondence and periods of half-integral weight modular forms, which seem to give stronger results in our setting.

### 1.2. Context

Let  $\mathcal{F}$  traverse a sequence of finite families of cusp forms defined on congruence covers of the modular surface  $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$  (examples will follow shortly). For each  $\varphi \in \mathcal{F}$ , we may define a probability measure

$\mu_\varphi$  on  $SL_2(\mathbb{Z})\backslash\mathbb{H}$  by pushforward of  $L^2$ -mass: for  $\Psi \in C_c(SL_2(\mathbb{Z})\backslash\mathbb{H})$ ,

$$\mu_\varphi(\Psi) := \frac{\langle \Psi\varphi, \varphi \rangle}{\langle \varphi, \varphi \rangle},$$

with  $\langle, \rangle$  the Petersson inner product. For the sequences of families considered in this article, one either knows or can show readily that the mean of the  $\mu_\varphi$  tends to the uniform measure as  $\mathcal{F}$  varies: for fixed  $\Psi$ ,

$$\frac{1}{|\mathcal{F}|} \sum_{\varphi \in \mathcal{F}} \mu_\varphi(\Psi) = \frac{\langle \Psi, 1 \rangle}{\langle 1, 1 \rangle} + o(1).$$

We consider here the problem of bounding or estimating the variance sums

$$V_{\mathcal{F}}(\Psi) := \sum_{\varphi \in \mathcal{F}} |\mu_\varphi(\Psi)|^2 \tag{1.1}$$

for nice enough fixed  $\Psi : SL_2(\mathbb{Z})\backslash\mathbb{H} \rightarrow \mathbb{C}$  of mean zero. For concreteness and simplicity, we suppose throughout this article that  $\Psi$  is a fixed even Hecke–Maass cusp form, noting that some results quoted below apply to more general observables than this.

The problem of estimating  $V_{\mathcal{F}}(\Psi)$  becomes more difficult the smaller the family  $\mathcal{F}$  is relative to the parameters of its typical elements. To illustrate, let  $1 \leq H \leq T$ . Let  $\mathcal{F}([T, T + H])$  denote the set of normalized cuspidal Hecke–Maass forms on  $SL_2(\mathbb{Z})\backslash\mathbb{H}$  of eigenvalue  $1/4 + t^2$  for some  $t \in [T, T + H]$ . Here,  $H$  varies with  $T$  as  $T \rightarrow \infty$ . One knows that  $\#\mathcal{F}([T, T + H]) \asymp TH$ . The general quantum ergodicity theorem implies (for the analogous problem on much more general manifolds) that

$$V_{\mathcal{F}([T, T + H])} \ll \frac{T^2}{\log T},$$

but one expects the much stronger upper bound

$$V_{\mathcal{F}([T, T + H])}(\Psi) \ll_\varepsilon HT^\varepsilon, \tag{1.2}$$

which should moreover be essentially sharp (that is, up to the factor  $T^\varepsilon$ ). This expectation is a consequence of the Lindelöf hypothesis combined with the triple product formula in the form

$$|\mu_\varphi(\Psi)|^2 = T^{-1+o(1)} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) \tag{1.3}$$

for  $\varphi \in \mathcal{F}([T, T + H])$ , where  $o(1)$  denotes a quantity tending to zero with  $T$ . In the ‘long family’ case  $H = T$ , an asymptotic formula for  $V_{\mathcal{F}([T, 2T])}(\Psi)$  (confirming a more precise version of (1.2)) follows from work of P. Zhao [45]. Jung [17] has confirmed the expectation (1.2) for  $H \geq T^{1/3+\varepsilon}$ , which appears to be the limit of current technology. The upper bound

$$V_{\mathcal{F}([T, T+1])}(\Psi) \ll_\varepsilon T^{1/3+\varepsilon} \tag{1.4}$$

obtained from the case  $H = T^{1/3+\varepsilon'}$  of (1.2) by positivity likewise appears to be the best to hope for in the near future. Results concerning holomorphic forms entirely analogous to those quoted above had been obtained earlier in a series of papers by Luo–Sarnak [21, 22, 23].

### 1.3. Main result

We pursue here level aspect analogues of the estimate (1.4): instead of working with increasing eigenvalues on a fixed surface, we consider bounded eigenvalues on a tower of congruence covers of the modular surface. To that end, fix  $\Lambda > 1/4$ . Let  $p$  denote a large prime, regarded as tending to  $\infty$ . Let  $\mathcal{F}(p)$  denote the set of normalized Hecke–Maass cuspidal newforms  $\varphi$  on  $\Gamma_0(p)$  whose Laplace

eigenvalue is at most  $\Lambda$ . One knows that  $\#\mathcal{F}(p) \asymp p$  (see [16, Section 15.5]), so the trivial estimate is  $V_{\mathcal{F}(p)}(\Psi) \ll p$ . The triple product formula, as specialized in [25, Section 4], gives the identity

$$|\mu_\varphi(\Psi)|^2 = p^{-1+o(1)} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}), \tag{1.5}$$

where  $o(1)$  denotes a quantity tending to zero with  $p$ . The Lindelöf hypothesis thus suggests that

$$\mu_\varphi(\Psi) \ll p^{-1/2+o(1)} \tag{1.6}$$

for individual  $\varphi \in \mathcal{F}(p)$ , and hence that

$$V_{\mathcal{F}(p)}(\Psi) \ll_\varepsilon p^\varepsilon. \tag{1.7}$$

The nontrivial estimate  $V_{\mathcal{F}(p)}(\Psi) \ll p/\log(p)$  likely follows, in a more general setting, from the methods of [19] (generalized to non-compact quotients, and using [33] to verify the hypothesis of Benjamini–Schramm convergence). We establish the following further strengthening:

**Theorem 1.** *Fix an even Hecke–Maass cusp form  $\Psi$  on  $SL_2(\mathbb{Z})$  and  $\Lambda > 1/4$ . Let  $\mathcal{F}(p)$  be as above, and  $V_{\mathcal{F}(p)}$  as in (1.1). Then*

$$V_{\mathcal{F}(p)}(\Psi) \ll_\varepsilon p^{1/4+\varepsilon}. \tag{1.8}$$

By Chebyshev’s inequality, we deduce the following approximation to (1.6):

**Theorem 2.** *Fix positive reals  $\alpha, \beta$ , for which  $2\alpha + \beta < 3/4$ . Then*

$$\#\{\varphi \in \mathcal{F}(p) : |\mu_\varphi(\Psi)| > p^{-\alpha}\} \ll p^{-\beta} \#\mathcal{F}(p).$$

### 1.4. Conditional sharp bounds

While the Lindelöf-consistent conjecture (1.7) appears to be out of reach, we give a conditional proof that appears to be the first of its kind. Here and henceforth, let  $\chi_d$  denote the quadratic Dirichlet character attached to a fundamental discriminant  $d$ .

**Theorem 3.** *Assume that*

1. *The Lindelöf hypothesis  $L(\Psi \otimes \chi_d, \frac{1}{2}) \ll_\varepsilon d^\varepsilon$  holds for the family of quadratic twists of  $\Psi$ , and that*
2. *The Ramanujan conjecture holds for the Hecke eigenvalues of  $\Psi$ .*

*Then (1.7) holds.*

### 1.5. Application to moments of triple product L-functions

Under (1.5), Theorems 1 and 3 translate to moment bounds for the (non-negative) central values of some triple product  $L$ -functions:

**Theorem 4.** *Unconditionally,*

$$\sum_{\varphi \in \mathcal{F}(p)} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) \ll p^{5/4+o(1)}. \tag{1.9}$$

*Under the assumptions of Theorem 3,*

$$\sum_{\varphi \in \mathcal{F}(p)} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) \ll p^{1+o(1)}. \tag{1.10}$$

By comparison, Jung’s estimate (1.4) translates to

$$\sum_{\varphi \in \mathcal{F}([T, T+1])} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) \ll_{\varepsilon} T^{4/3+\varepsilon}. \tag{1.11}$$

The first moments on the LHS of (1.9) and (1.11) are analogous in that in each, the analytic conductor ( $\asymp p^4$  and  $\asymp T^4$ ) is roughly the fourth power of the family cardinality ( $\asymp p$  and  $\asymp T$ ). We note also that Iwaniec–Michel [13] established (the analogue for holomorphic forms of) the estimate

$$\sum_{\varphi \in \mathcal{F}(p)} L(\varphi \otimes \varphi, \frac{1}{2})^2 \ll p^{1+o(1)}, \tag{1.12}$$

which may be understood as a variant of (1.10), in which  $\Psi$  is an Eisenstein series.

**Remark 1.** A general rule of thumb in the literature on moment bounds for families of  $L$ -functions is that one should be able to establish Lindelöf-consistent bounds when the family size is at least the fourth root of the analytic conductor. This rule does not seem to apply when one considers ‘sparse’ moments, such as (1.9); one may understand sparsity here as coming from the coincidence of two of the three factors in the triple product  $L$ -parameter coincide. By comparison, it is not difficult to prove that

$$\sum_{\varphi \in \mathcal{F}(p)} L(\varphi \otimes \Psi_1 \otimes \Psi_2, \frac{1}{2}) \ll p^{1+o(1)}$$

for fixed  $\Psi_1, \Psi_2$  (see [9] for exact formulas of a similar spirit).

**Remark 2.** We were unable to obtain a ‘classical’ proof of (1.9) using the approximate functional equation and familiar transformations thereafter.

**1.6. Application to sparse moments of quadratic twists**

The following curious bound is a byproduct of our method:

**Theorem 5.** For  $C \geq 1$ , one has

$$\sum_{n: |n| \leq Cp} L(\Psi \otimes \chi_{pn}, \frac{1}{2}) \left( 1 + \log \left( \frac{Cp}{|n|} \right) \right) \ll Cp \tag{1.13}$$

uniformly in  $C, p$ , where the sum is over integers  $n$  for which  $pn$  is a fundamental discriminant. In particular,

$$\sum_{n: |n| < p} L(\Psi \otimes \chi_{pn}, \frac{1}{2}) \ll p. \tag{1.14}$$

In fact, the methods of this paper reveal a surprising relationship between the moments  $\sum_n L(\Psi \otimes \chi_{pn}, \frac{1}{2})$  (possibly over shorter intervals than those above) and  $\sum_{\varphi \in \mathcal{F}(p)} L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2})$ . This relationship demonstrates the difficulty underlying an unconditional proof of (1.7). We refer to Section 2 for a detailed discussion of this relationship but record here one consequence:

**Theorem 6.** Assume (1.7), or equivalently, (1.10). Assume also that  $L(\Psi, \frac{1}{2}) \neq 0$ . Then

$$\sum_{n: |n| < p} \frac{L(\Psi \otimes \chi_{pn}, \frac{1}{2})}{|n|^{1/2}} \ll p^{1/2+o(1)}. \tag{1.15}$$

The Lindelöf hypothesis implies that the conclusion of Theorem 1.15 holds without the assumption that  $L(\Psi, \frac{1}{2}) \neq 0$ , but this assumption is necessary for our proof. The issue is that if  $L(\Psi, \frac{1}{2}) = 0$ , then

$L(\varphi \otimes \varphi \otimes \Psi, \frac{1}{2}) = L(\text{ad } \varphi \otimes \Psi, \frac{1}{2})L(\Psi, \frac{1}{2}) = 0$ , so our hypotheses (1.7) and (1.10) hold trivially; hence their assumption is not helpful.

We note that the family size in (1.14) is  $\asymp p$ , while the analytic conductor in the largest dyadic range of the sum is  $\asymp p^4$ , so one might expect the difficulty of the moment problem addressed by (1.14) to be comparable to that for

$$\sum_{n:|n|<p} L(\Psi \otimes \chi_n, \frac{1}{2})^2, \tag{1.16}$$

$$\sum_{n:|n|<p} L(\chi_{pn}, \frac{1}{2})^2, \tag{1.17}$$

or

$$\sum_{n:|n|<p} L(\chi_n, \frac{1}{2})^4. \tag{1.18}$$

One may understand (1.17) as the variant of (1.14) obtained by taking for  $\Psi$  an Eisenstein series. Heath–Brown [10] proved an upper bound  $p^{1+o(1)}$  for (1.18), and it seems likely that the same proof works also for (1.17) and (1.16); a closely related argument appears implicitly in [13]. We note also that Soundararajan–Young [39] have established an asymptotic formula, conditional on GRH, for a mild variant of (1.16).

The unconditional estimate (1.14) established here seems beyond the limits the methods indicated in the preceding paragraph: after applying an approximate functional equation, one faces (smooth) sums roughly of the shape

$$S := \sum_{m \sim p^2, n \sim p} \frac{\lambda(m)\chi_{pn}(m)}{\sqrt{m}}, \tag{1.19}$$

where  $\lambda(m)$  denotes the  $m$ th normalized Fourier coefficient of  $\Psi$ . To establish (1.14) in this way, one must show that  $S \ll p^{1+o(1)}$ , which seems hopeless.

We note that the proof of (1.14) is specific to the central point  $s = 1/2$ , while the proofs indicated above of analogous estimates for (1.16), (1.17), or (1.18) apply more generally to  $s = 1/2 + it$  for any fixed  $t$ .

We prove Theorem 5 at the end of Section 4. The proof goes by the connection between the values  $L(\Psi \otimes \chi_{pn}, \frac{1}{2})$  and the squared magnitudes  $|b(pn)|^2$  of the Fourier coefficients of a half-integral weight lift of  $\Psi$ ; what we really show is (for instance)

$$\sum_{n:|n|<p} |b(pn)|^2 \ll p. \tag{1.20}$$

Such estimates improve in the indicated range on the diagonal case of those of Duke–Iwaniec [8, (4)], which specialize to (the analogue for holomorphic forms of)

$$\sum_{n:|n|<p} |b(pn)|^2 \ll p^{3/2+\varepsilon}. \tag{1.21}$$

### 1.7. Method

The basic idea behind the proof is to write

$$\mu_\varphi(\Psi) = \langle \varphi, G \rangle \tag{1.22}$$

for some automorphic function  $G$  and then to estimate the  $L^2$ -norm  $\langle G, G \rangle$ . The method also applies to cocompact arithmetic hyperbolic surfaces. A high-level overview is given in Section 2. The overall

strategy is related to that employed in our work on the quantum variance [27, 29, 30] and subconvexity [31] problems, and also to recent work of Raphael Steiner [40] on the sup norm problem.

**2. Division of the proof**

The purpose of this section is to reduce the proof of our main results to that of some independent claims to be verified in the body of the paper.

**2.1. Jacobi theta function**

For  $z = x + iy \in \mathbb{H}$ , set  $\theta(z) := y^{1/4} \sum_{n \in \mathbb{Z}} e(n^2 z)$ , where  $e(z) := e^{2\pi iz}$ . By considering Fourier expansions at the cusps of  $\Gamma_0(4)$ , we obtain the crude upper bound

$$\theta(z) \ll \text{ht}(z)^{1/4}, \tag{2.1}$$

where  $\text{ht}(z) := \max_{\gamma \in \text{SL}_2(\mathbb{Z})} \text{Im}(\gamma z)$ .

**2.2. Theta multiplier**

For  $\gamma \in \Gamma_0(4)$ , set  $J(\gamma, z) := \theta(\gamma z)/\theta(z)$ . We recall from [37, Proposition 2.2] that if  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , then

$$J(\gamma, z) = \varepsilon(\gamma) \frac{\sqrt{cz + d}}{|cz + d|^{1/2}}, \quad \varepsilon(\gamma) = \varepsilon_d^{-1} \left( \frac{c}{d} \right). \tag{2.2}$$

Here we define the square root by  $\sqrt{z} := |z|^{1/2} e^{i \arg(z)/2}$  if  $z = |z|e^{i \arg(z)}$  with  $\arg(z) \in (-\pi, \pi]$ ,  $\varepsilon_d = 1$  or  $i$  according as  $d \equiv 1$  or  $-1$  modulo 4, and  $\left(\frac{c}{d}\right)$  as the ‘quadratic residue symbol’ characterized in [37, page 442].

**2.3. Petersson inner product**

For a congruence subgroup  $\Gamma \leq \Gamma_0(4)$  and  $\kappa \in \mathbb{Z}$ , we call  $F : \mathbb{H} \rightarrow \mathbb{C}$  modular of weight  $\kappa/2$  on  $\Gamma$  if  $F(\gamma z) = J(\gamma, z)^\kappa F(z)$  for  $\gamma \in \Gamma$ . If  $F_1, F_2$  are modular of weight  $\kappa/2$  on  $\Gamma$ , then the function  $F_1 \overline{F_2}$  is  $\Gamma$ -invariant; if it induces a function on  $\Gamma \backslash \mathbb{H}$  that is integrable with respect to the measure  $d\mu(z) := \frac{dx dy}{y^2}$ , then we define the normalized Petersson inner product

$$\langle F_1, F_2 \rangle := \frac{1}{[\text{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]} \int_{z \in \Gamma \backslash \mathbb{H}} F_1(z) \overline{F_2(z)} d\mu(z)$$

and associated norm  $\|F\| := \langle F, F \rangle^{1/2}$ ; here,  $\overline{\Gamma} \leq \text{PSL}_2(\mathbb{Z})$  denotes the image of  $\Gamma$ . Note that the definition of  $\langle, \rangle$  is invariant under shrinking  $\Gamma$ .

**2.4. The half-integral weight lift and its Fourier expansion**

We recall in Section 3.3.2 the construction via theta lifting of a Maass/inverse-Shimura/Shintani lift  $h$  of  $\Psi$ . It is nonzero precisely when  $L(\Psi, \frac{1}{2}) \neq 0$ . It is modular of weight  $1/2$  on  $\Gamma_0(4)$  and belongs to an analogue of the Kohnen plus space. It admits a Fourier expansion

$$h(z) = \sum_{n \in \mathbb{Z}, n \neq 0} \frac{b(n)}{|n|^{1/2}} W(ny) e(nx)$$

where  $W$  is a Whittaker function (see Section 3.3.3) and  $b(n) = 0$  unless  $n \equiv 0, 1 \pmod{4}$ . When  $d$  is a fundamental discriminant, we have by [7, (5.17)]

$$|b(d)|^2 = cL(\Psi \otimes \chi_d, \frac{1}{2}). \tag{2.3}$$

Here  $c$  depends only upon  $\Psi$ , and  $c \neq 0$  whenever  $h \neq 0$ . More generally, any  $n \equiv 0, 1 \pmod{4}$  may be written uniquely as  $n = d\delta^2$ , where  $d$  is a fundamental discriminant and  $\delta$  is a natural number; we may then deduce via the Shimura relation the estimate (see [20, Proposition 6.1])

$$|b(n)| \ll_{\varepsilon} |b(d)|\delta^{\vartheta+\varepsilon},$$

where  $\vartheta \in [0, 7/64]$  quantifies the known bounds towards the Ramanujan conjecture for the Hecke eigenvalues of  $\Psi$ . Conrey–Iwaniec [5] have shown that  $L(\Psi \otimes \chi_d, \frac{1}{2}) \ll_{\varepsilon} d^{1/3+\varepsilon}$ . Since  $\vartheta \leq 1/3$ , it follows in general that

$$|b(n)|^2 \ll_{\varepsilon} |n|^{1/3+\varepsilon}. \tag{2.4}$$

On the other hand, it is expected (by the Lindelöf hypothesis for  $L(\Psi \otimes \chi_d, \frac{1}{2})$  and the Ramanujan conjecture for  $\Psi$ ) that

$$|b(n)|^2 \ll_{\varepsilon} |n|^{\varepsilon}. \tag{2.5}$$

### 2.5. Application of an incomplete Hecke operator

We denote by  $h^{\sharp}$  the (normalized) application to  $h$  of a variant of the classical ‘ $U_p$ ’ operator:

$$h^{\sharp}(z) := \frac{1}{p^{1/2}} \sum_{j \in \mathbb{Z}/p\mathbb{Z}} h\left(\frac{z + pj}{p^2}\right) = \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{b(pn)}{|n|^{1/2}} W\left(\frac{ny}{p}\right) e\left(\frac{nx}{p}\right). \tag{2.6}$$

We record in Section 3.4.5 the level of  $h^{\sharp}$ . We show in Section 4 that

$$\langle h^{\sharp}, h^{\sharp} \rangle = \langle h, h \rangle, \tag{2.7}$$

which reflects a special feature of  $1/2$ -integral weight forms.

### 2.6. Properties of the varying forms

We assume that each  $\varphi \in \mathcal{F}(p)$  is arithmetically normalized, so that

$$\varphi(z) = \sum_{n \in \mathbb{Z}_{\neq 0}} \frac{\lambda_{\varphi}(|n|)}{|n|^{1/2}} W_{\varphi}(ny) e(nx),$$

where  $\lambda_{\varphi}(1) = 1$  and  $W_{\varphi}(y) = 2y^{1/2} K_{it_{\varphi}}(2\pi y)$  for  $y > 0$ , where  $1/4 + t_{\varphi}^2$  is the Laplace eigenvalue of  $\varphi$ . Each  $\varphi \in \mathcal{F}(p)$  is an eigenfunction of the Atkin–Lehner/Fricke involution with eigenvalue  $\pm 1$ ; that is to say,  $\varphi(-1/(pz)) = \pm\varphi(z)$ , or equivalently,

$$\varphi(-1/z) = \pm\varphi(z/p), \tag{2.8}$$

and it is known [11] that

$$p^{-\varepsilon} \ll_{\varepsilon} \langle \varphi, \varphi \rangle \ll_{\varepsilon} p^{\varepsilon}. \tag{2.9}$$



2.7. An explicit seesaw identity

We show in Section 5 that for  $\varphi \in \mathcal{F}(p)$ ,

$$\mu_\varphi(\Psi) = \pm 4 \langle \varphi(\frac{4z}{p}) \vartheta(z), h^\sharp(z) \rangle, \tag{2.10}$$

where  $\pm$  is as in (2.8). Here we abuse notation mildly by writing simply (for example)  $\varphi(\frac{4z}{p})$  for the function  $z \mapsto \varphi(\frac{4z}{p})$ . A notable feature of the RHS of (2.10) is that it depends linearly upon  $\varphi$ .

**Remark.** Related identities involving forms of level 1 have been given by Biro [2]; the proof given here is different and applies more generally (for example, also to compact arithmetic quotients  $\Gamma \backslash \mathbb{H}$ ). See also Ichino [12, Section 11]. Similar identities are also implicit in our work [27, 29, 30] on the quantum variance problem.

2.8. Reduction to period bounds

By (2.9), (2.10) and Bessel’s inequality, we have

$$\sum_{\varphi \in \mathcal{F}(p)} |\mu_\varphi(\Psi)|^2 = 16 \sum_{\varphi \in \mathcal{F}(p)} \frac{|\langle \varphi(\frac{4z}{p}), \overline{\theta(z)} h^\sharp(z) \rangle|^2}{\langle \varphi, \varphi \rangle} \ll p^{o(1)} \langle \overline{\theta} h^\sharp, \overline{\theta} h^\sharp \rangle, \tag{2.11}$$

so the proof of Theorem 1 reduces to that of the estimate

$$\|\theta h^\sharp\|^2 \ll p^{1/4+o(1)}. \tag{2.12}$$

**Remark.** This part of the argument is reminiscent of arguments in [3, 34, 1] and (implicitly) in [42, 24], among other places. A more general (but less elementary) approach to identities like (2.11) is given in [31], following [32].

2.9. The basic inequality

We show in Section 6 that for any  $T \geq 1$ ,

$$\|\theta h^\sharp\|^2 \ll T^{1/2} + p^{-1/2} R, \tag{2.13}$$

where

$$R := \sum_n \frac{|b(pn)|^2}{|n|^{1/2}} \left(1 + \frac{|n|}{p/T}\right)^{-100}.$$

This is a key step in the argument, so we sketch here the basic idea behind the proof. We estimate separately the contributions to  $\|\theta h^\sharp\|^2$  from the ranges  $\{z : \text{ht}(z) \leq T\}$  and  $\{z : \text{ht}(z) > T\}$ , where  $\text{ht}$  is as in Section 2.1. In both ranges, we apply the pointwise bound (2.1) for  $\theta$ . For the range where  $\text{ht}(z) \leq T$ , we apply the  $L^2$ -bound (2.7) for  $h^\sharp$  to obtain the estimate  $\ll T^{1/2}$ . For the range where  $\text{ht}(z) > T$ , we apply Parseval to the Fourier expansion of  $h^\sharp$  and appeal to the rapid decay of the Whittaker function  $W$  to establish estimates such as

$$\frac{1}{p} \int_{y=T}^\infty y^{1/2} \int_{x=0}^p |h^\sharp(x + iy)|^2 \frac{dx dy}{y^2} \ll p^{-1/2} R. \tag{2.14}$$

**Remark.** It would be possible to refine the present analysis by employing the spectral expansion of  $|\theta|^2$  as in [28] in place of the upper bound (2.1), but doing so does not seem to lead to stronger unconditional results.

**2.10. Completion of the proof**

By taking  $T = p^{1/2}$  in (2.13) and appealing to the Conrey–Iwaniec bound (2.4), we readily obtain (2.12), hence Theorem 1. For the proof of Theorem 3, we note that its hypotheses imply (2.5), which gives the required bound upon taking  $T = 1$  in (2.13).

The proof of Theorem 6 is recorded in Section 7.

**Remark 1.** Taking  $T = p^{1+\varepsilon}$  in (2.13) and appealing to any convexity bound of the form  $b(pn) \ll (pn)^{O(1)}$  already gives the nontrivial estimate  $\sum_{\varphi \in \mathcal{F}(p)} |\mu_\varphi(\Psi)|^2 \ll p^{1/2+o(1)}$ .

**Remark 2.** It is natural to ask the questions:

1. Is the Conrey–Iwaniec bound an essential input to the method?
2. Can one do better by exploiting the average over  $n$  in  $R$ ?

To address these, let  $R^b$  denote the subsum over  $R$  obtained by restricting to summation indices  $n$  for which  $pn$  is a fundamental discriminant. Then

$$R^b = c \sum_{\substack{n: \\ pn \text{ is fundamental}}} \frac{L(\Psi \otimes \chi_{pn}, \frac{1}{2})}{|n|^{1/2}} \left(1 + \frac{|n|}{p/T}\right)^{-100}$$

for some  $c$  depending only upon  $\Psi$ . It seems likely that

1. One can show directly using an approximate functional equation [16, Section 5.2] and Heath–Brown’s large sieve for quadratic characters [10] that for  $1 \leq T \leq p$ , one has  $R^b \ll T^{-1/2} p^{1+o(1)}$ , and that
2. one can establish the same bound for  $R$  by using the Rankin–Selberg upper bound  $\sum_{n \leq x} |\lambda_\Psi(n)|^2 \ll x$  to control the contribution from non-fundamental discriminants.

If so, then by taking  $T = p^{1/2}$ , one obtains a proof of Theorem 1 that does not rely upon the Conrey–Iwaniec bound. Conversely, we do not know how to do better by exploiting the average over  $n$  in  $R$ .

**3. Preliminaries**

**3.1. Generalities**

**3.1.1.**

Let  $B, C$  be coprime natural numbers. Set

$$\Gamma_0(C/B) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) : B \mid b, C \mid c \right\}.$$

When  $B = 1$ , this is the standard definition of  $\Gamma_0(C) = \Gamma_0(C/1)$ . In general,  $\Gamma_0(C/B)$  is a congruence subgroup of  $\text{SL}_2(\mathbb{Z})$  that is conjugate to  $\Gamma_0(BC)$ .

As motivation for the notation, note that if  $F : \mathbb{H} \rightarrow \mathbb{C}$  is  $\text{SL}_2(\mathbb{Z})$ -invariant, then the function  $z \mapsto F(zC/B)$  is  $\Gamma_0(C/B)$ -invariant.

**3.1.2.**

Let  $R := M_2(\mathbb{Z})$  denote the ring of  $2 \times 2$  integral matrices. Set  $S := \mathbb{Z} + 2R$ . We use a superscripted 0 to denote ‘traceless’ elements, so that, for instance,

$$M_2(\mathbb{R})^0 = \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} : a, b, c \in \mathbb{R} \right\},$$

$$S^0 = \left\{ \begin{pmatrix} a & 2b \\ 2c & -a \end{pmatrix} : a, b, c \in \mathbb{Z} \right\}.$$

Note that

$$S = \mathbb{Z} \oplus S^0 \tag{3.1}$$

**3.1.3.**

For natural numbers  $B, C$ , set

$$R(C/B) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Q}) : a, d \in \mathbb{Z}; b \in B^{-1}\mathbb{Z}, c \in C\mathbb{Z} \right\}.$$

It is a lattice in  $M_2(\mathbb{Q})$ . We abbreviate  $R(C/1) := R(C)$ . We note that  $R(C/B)$  is not directly related to  $\Gamma_0(C/B)$  except when  $B = 1$ , in which case  $\Gamma_0(C) = \text{SL}_2(\mathbb{Z}) \cap R(C)$ . The significance of the notation is that  $R(C/B)$  and  $R(B/C)$  are dual lattices with respect to the quadratic form on  $M_2(\mathbb{Q})$  defined by the determinant.

When  $B, C$  are odd, we set  $S(C/B) := S \cap R(C/B) = \mathbb{Z} + 2R(C/B)$  and  $S^0(C/B) := S^0 \cap S(C/B)$ .

**3.1.4.**

For  $w = u + iv \in \mathbb{H}$ , define  $\sigma_w \in \text{SL}_2(\mathbb{R})$  by the formula

$$\sigma_w := \begin{pmatrix} v^{1/2} & uv^{-1/2} \\ & v^{-1/2} \end{pmatrix},$$

so that  $\sigma_w i = w$ .

**3.2. Theta kernels**

We recall the definitions and basic properties of some theta kernels. We refer to [38], [43, Section 2], and [26, Appendix B] for details.

In what follows, take  $w, w_1, w_2, z = x + iy \in \mathbb{H}$ .

**3.2.1.**

Define  $P : M_2(\mathbb{R}) \rightarrow \mathbb{R}$  by

$$P \left( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) := \frac{a^2 + b^2 + c^2 + d^2}{2},$$

$\phi_{w,z}^0 : M_2(\mathbb{R})^0 \rightarrow \mathbb{C}$  by

$$\phi_{w,z}^0(\alpha) := \frac{1}{2\pi} y^{3/4} \exp(-2\pi y P(\sigma_w^{-1} \alpha \sigma_w)) e(x \det(\alpha)),$$

and  $\phi_{w_1, w_2, z} : M_2(\mathbb{R}) \rightarrow \mathbb{C}$  by

$$\phi_{w_1, w_2, z}(\alpha) := \frac{1}{2\pi} y \exp(-2\pi y P(\sigma_{w_1}^{-1} \alpha \sigma_{w_2})) e(x \det(\alpha)).$$

Note that for  $\alpha = m + \beta$  with  $m \in \mathbb{R}, \beta \in M_2(\mathbb{R})^0$ ,

$$\phi_{w, w, z}(\alpha) = y^{1/4} e(m^2 z) \phi_{w, z}^0(\beta). \tag{3.2}$$

**3.2.2.**

Set

$$\theta(w, z) := \sum_{\alpha \in S^0} \phi_{w, z}^0(\alpha).$$

Then  $\theta$  defines a theta kernel (à la Maass–Shintani–Waldspurger), of weight  $-1/2$  on  $\Gamma_0(4)$  in the variable  $w$  and of weight  $0$  on  $\text{SL}_2(\mathbb{Z})$  in the variable  $z$  (see also [38], [18, Section 2]).

**3.2.3.**

For a lattice  $L \subseteq M_2(\mathbb{Q})$ , set

$$\theta(L; w_1, w_2, z) := \sum_{\alpha \in L} \phi_{w_1, w_2, z}(\alpha).$$

This defines a modular function of weight 0 in each variable with respect to suitable congruence subgroups (see [38]).

**3.2.4.**

The lattice  $R(p)$  has discriminant  $p^2$  and dual  $R(1/p)$ . The quadratic form  $(R(p), \det)$  has signature  $(2, 2)$ . Thus (see [38])

$$\theta(R(p); w_1, w_2, -1/z) = p^{-1}\theta(R(1/p); w_1, w_2, z). \tag{3.3}$$

**3.3. Ternary theta lifts**

**3.3.1.**

We assume that  $\Psi : \text{SL}_2(\mathbb{Z}) \backslash \mathbb{H} \rightarrow \mathbb{C}$  is arithmetically normalized, so that its Fourier expansion reads

$$\Psi(w) = \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{\lambda_\Psi(n)}{|n|^{1/2}} W_\Psi(ny) e(nx)$$

where  $\lambda_\Psi(n) = \lambda_\Psi(|n|)$  satisfies  $\lambda_\Psi(1) = 1$  and so that the Ramanujan conjecture reads  $|\lambda_\Psi(n)| \leq \sum_{d|n} 1$ , while  $W_\Psi(y) = 2|y|^{1/2} K_{it_\Psi}(2\pi|y|)$ , with  $1/4 + t_\Psi^2$  the Laplace eigenvalue of  $\Psi$ .

**3.3.2.**

Define  $h : \mathbb{H} \rightarrow \mathbb{C}$  by requiring that

$$\overline{h(z)} := \int_{w \in \text{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} \theta(w, z) \Psi(w) d\mu(w).$$

Then  $h$  is a constant multiple of the form constructed in [18, Proposition 2.3] and has the properties indicated in Section 2.4 (compare with [7, Section 5] and [20, Section 6]).

**3.3.3.**

The Whittaker function  $W$  of  $h$  is given by  $W(y) = W_{\text{sgn}(y)/4, it_\Psi/2}(4\pi|y|)$ . Since  $\Psi$  is non-constant, we have  $t_\Psi \in \mathbb{R}$  or  $it_\Psi \in (-1/2, 0) \cup (0, 1/2)$ . (The second possibility does not actually occur, thanks to the known Selberg eigenvalue conjecture for  $\text{PGL}_2(\mathbb{Z})$ , but we do not need to exclude it.) We require the estimate (see [41, Section 7.3], [18, (0.11)], [6, 13.14.21])

$$W(y) \ll \min(|y|^{1/2-\vartheta/2}, |y|^{1/4} e^{-2\pi|y|}). \tag{3.4}$$

Here the implied constant and  $\vartheta \in (|\text{Im}(t_\Psi)|, 1/2)$  depend upon  $t_\Psi$ , which is fixed for us. In particular,  $W(y) \ll |y|^{1/4+\delta}$  for some fixed  $\delta > 0$ .

**3.3.4.**

A set of inequivalent cusps for  $\Gamma_0(4)$  is given by  $\{\infty, 0, 1/2\}$ . It is shown in [20, Section 11] that

$$e^{\pi i/4} (z/|z|)^{-1/2} h(-1/4z) = \sqrt{2} \sum_{n \equiv 0(4)} \frac{b(n)}{|n|^{1/2}} W\left(\frac{ny}{4}\right) e\left(\frac{nx}{4}\right) \tag{3.5}$$

and

$$e^{\pi i/4}(z/|z|)^{-1/2}h\left(\frac{1}{2} - \frac{1}{4z}\right) = \sqrt{2} \sum_{n \equiv 1(4)} \frac{b(n)}{|n|^{1/2}} W\left(\frac{ny}{4}\right) e\left(\frac{nx}{4}\right). \tag{3.6}$$

Thus the expansion of  $h$  at any cusp of  $\Gamma_0(4)$  is obtained from that at the cusp  $\infty$ , essentially by restricting the summation index to some congruence class modulo 4. (Strictly speaking, since  $h$  has half-integral weight, ‘the’ expansion of  $h$  at a cusp depends upon the choice of matrix to represent the cusp. We hope this abuse of terminology introduces no confusion.)

For  $\ell = 1, 2$ , set

$$h_\ell(z) := \left(\frac{\ell z + 1}{|\ell z + 1|}\right)^{-1/2} h\left(\begin{pmatrix} 1 & \\ \ell & 1 \end{pmatrix} z\right).$$

Using the readily verified identities

$$h\left(\begin{pmatrix} 1 & \\ 1 & 1 \end{pmatrix} z\right) = h\left(\frac{-1}{z+1}\right), \quad h\left(\begin{pmatrix} 1 & \\ 2 & 1 \end{pmatrix} z\right) = h\left(\frac{1}{2} - \frac{1/2}{2z+1}\right),$$

we may read off the Fourier expansions of  $h_1$  and  $h_2$  at the cusp  $\infty$  from (3.5) and (3.6).

### 3.4. Cusps of $\Gamma_0(4/p)$

For general background on cusps and fundamental domains, we refer to [36, 14, 15].

#### 3.4.1.

For the remainder of Section 3.4, we set  $\Gamma := \Gamma_0(4/p)$ . We note that  $-1 \in \Gamma$ .

#### 3.4.2.

Set  $\Delta := \left\{ \pm \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} : n \in \mathbb{Z} \right\} \leq \text{SL}_2(\mathbb{Z})$ . Then any set  $\mathcal{C} := \{\gamma_1, \dots, \gamma_6\} \subseteq \text{SL}_2(\mathbb{Z})$  consisting of elements of the form

$$\begin{aligned} \gamma_1 &:= \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, & \gamma_2 &:= \begin{pmatrix} 1 & \\ 2 & 1 \end{pmatrix}, & \gamma_3 &:= \begin{pmatrix} 1 & \\ 1 & 1 \end{pmatrix}, \\ \gamma_4 &:= \begin{pmatrix} p & * \\ 4 & * \end{pmatrix}, & \gamma_5 &:= \begin{pmatrix} p & * \\ 2 & * \end{pmatrix}, & \gamma_6 &:= \begin{pmatrix} p & * \\ 1 & * \end{pmatrix} \end{aligned}$$

gives representatives for the double coset space  $\Gamma \backslash \text{SL}_2(\mathbb{Z}) / \Delta$ , and  $\{\gamma_j \infty : j = 1, \dots, 6\} = \{\infty, 1/2, 1, p/4, p/2, p\}$  is a maximal set of inequivalent cusps for  $\Gamma \backslash \mathbb{H}$ .

#### 3.4.3.

For  $\gamma \in \text{SL}_2(\mathbb{Z})$ , the width of the cusp  $\gamma \infty$  for  $\Gamma \backslash \mathbb{H}$  is the cardinality  $w(\gamma)$  of the preimage in  $\Gamma \backslash \text{SL}_2(\mathbb{Z})$  of  $\Gamma \gamma \Delta$ . We have

$$w(\gamma_j) = p, p, 4p, 1, 1, 4 \text{ for } j = 1, \dots, 6, \text{ respectively.}$$

#### 3.4.4.

Recall that  $\text{ht} : \Gamma \backslash \mathbb{H} \rightarrow \mathbb{C}$  is defined by  $\text{ht}(z) := \max_{\gamma \in \text{SL}_2(\mathbb{Z})} \text{Im}(\gamma z)$ . By tiling  $\Gamma \backslash \mathbb{H}$  by translates of the standard fundamental domain for  $\text{SL}_2(\mathbb{Z})$ , we see that  $\text{ht}(z) \geq \sqrt{3}/2$  for all  $z$ . Moreover, for any  $T > 1$ , the union

$$\cup_{\gamma \in \mathcal{C}} \{\gamma(x + iy) : 0 \leq x \leq w(\gamma), y \geq T\} \tag{3.7}$$

is essentially disjoint and gives a fundamental domain for  $\{z \in \Gamma \backslash \mathbb{H} : \text{ht}(z) \geq T\}$ . Finally, for given  $y_0 \in (0, 1)$ , the fibers of the natural map

$$\{x + iy : 0 \leq x \leq p, y \geq y_0\} \rightarrow \Gamma \backslash \mathbb{H}$$

have cardinality  $O(1/y_0)$ , uniformly in  $p$  [15, Lemma 2.10].

**3.4.5.**

Using [37, Propositions 1.3, 1.5] (or, more precisely, their analogue for Maass forms), we verify that  $h^\sharp$  is modular (of weight  $1/2$ ) on  $\Gamma$ . Using §3.3.4, we see that the Fourier expansion of  $h^\sharp$  at the cusps  $1/2, 1$  of  $\Gamma$  is obtained from its expansion (2.6) at  $\infty$ , essentially by restricting the summation index to a congruence class modulo 4. More precisely, for  $\ell = 1, 2$ , define  $h_\ell^\sharp$  in terms of  $h_\ell$  analogously to how  $h^\sharp$  was defined in terms of  $h$ .

**Lemma 1.** *We have*

$$\left(\frac{\ell z + 1}{|\ell z + 1|}\right)^{-1/2} h^\sharp\left(\begin{pmatrix} 1 & \\ \ell & 1 \end{pmatrix} z\right) = h_\ell^\sharp(z). \tag{3.8}$$

*Proof.* Set  $w := \begin{pmatrix} 1 & \\ \ell & 1 \end{pmatrix} z$ . It suffices to show for each  $j \in \mathbb{Z}/p\mathbb{Z}$  that

$$\left(\frac{\ell z + 1}{|\ell z + 1|}\right)^{-1/2} h\left(\frac{w + pj}{p^2}\right) = h_\ell\left(\frac{z + pj}{p^2}\right).$$

We may assume that  $j$  is represented by a negative integer divisible by 4. The conclusion follows then in a straightforward manner from the second assertion of [37, Lemma 3.4] with  $N := 4, M := 1, K := 4$ , and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} p^{-1} & j \\ & p \end{pmatrix} \begin{pmatrix} 1 & \\ \ell & 1 \end{pmatrix} \begin{pmatrix} p^{-1} & j \\ & p \end{pmatrix}^{-1} = \begin{pmatrix} 1 + jp\ell & -j^2\ell \\ p^2\ell & 1 - jp\ell \end{pmatrix}.$$

□

**3.5. Explicit Shimizu lifts**

We make use of the following explicit form of the Shimizu correspondence.

**Lemma 2.** *Let  $\varphi \in \mathcal{F}(p)$ . Then for  $w_1, w_2, z \in \mathbb{H}$ ,*

$$\overline{\varphi(w_1)}\varphi(w_2) = \int_{z \in \Gamma_0(p) \backslash \mathbb{H}} \theta(R(p); w_1, w_2, z)\varphi(z) d\mu(z). \tag{3.9}$$

*Proof.* [26, Theorem 5.2] implies an analogous assertion for holomorphic forms. Running through the proof of that theorem with ‘ $k := 0$ ’, we obtain the identity stated here. □

**4. The  $L^2$ -norm of  $h^\sharp$**

In this section, we establish (2.7). We open  $\langle h^\sharp, h^\sharp \rangle$  as a double sum over  $j_1, j_2$ . The diagonal  $j_1 = j_2$  contributes  $\langle h, h \rangle$ ; our task is thus to show that

$$\sum_{j_1, j_2 \in \mathbb{Z}/p:} \left\langle h\left(\frac{z + pj_1}{p^2}\right), h\left(\frac{z + pj_2}{p^2}\right) \right\rangle = 0, \tag{4.1}$$

where here and henceforth we mildly abuse notation by writing (for example)  $h\left(\frac{z+pji}{p^2}\right)$  for the function  $z \mapsto h\left(\frac{z+pji}{p^2}\right)$ . By a simple change of variables, our task (4.1) reduces to verifying that

$$\sum_{j=1}^{p-1} \langle h(z), h(z + j/p) \rangle = 0. \tag{4.2}$$

To that end, temporarily fix  $j \in \{1, \dots, p - 1\}$ . Choose  $\gamma = \begin{pmatrix} a & * \\ pc & d \end{pmatrix} \in \Gamma_0(4)$  with  $cj \equiv -1 \pmod{p}$ ,  $c > 0$  and  $d \equiv 1 \pmod{4p}$ , so that also  $a \equiv 1 \pmod{4p}$ . Set

$$n(j/p) := \begin{pmatrix} 1 & j/p \\ & 1 \end{pmatrix}, \quad t(p^{-1}) := \begin{pmatrix} p^{-1} & \\ & p \end{pmatrix}$$

and  $\delta := n(j/p)\gamma t(p^{-1})$ . By direct calculation, we see that  $\delta$  belongs to  $\Gamma_0(4)$  and has the form  $\begin{pmatrix} * & * \\ c & pd \end{pmatrix}$ . By the invariance of the Petersson inner product and the modularity of  $h$ , we deduce that

$$\begin{aligned} \langle h(z), h(z + j/p) \rangle &= \langle h(z), h(n(j/p)z) \rangle \\ &= \langle h(\gamma z), h(n(j/p)\gamma z) \rangle \\ &= \langle h(\gamma z), h(\delta p^2 z) \rangle \\ &= \eta(j) \langle h(z), h(p^2 z) \rangle \end{aligned}$$

where

$$\begin{aligned} \eta(j) &:= \exp\left(\frac{i}{2} \underbrace{(\arg(pc z + d) - \arg(cp^2 z + pd))}_{=0}\right) \varepsilon(\gamma) \overline{\varepsilon(\delta)} \\ &= \varepsilon_d^{-1} \varepsilon_{pd} \left(\frac{pc}{d}\right) \left(\frac{c}{pd}\right) = \varepsilon_p \left(\frac{-j}{p}\right). \end{aligned}$$

In the final step, we invoked our assumptions on  $c, d$  and the rules of [37, page 442]. Thus  $(\mathbb{Z}/p\mathbb{Z})^\times \ni j \mapsto \eta(j)$  defines a constant multiple of the nontrivial quadratic character, and hence its sum over  $j = 1, \dots, p - 1$  vanishes. This completes the proof of (4.2) and hence of (4.1), and hence of (2.7).

**Remark.** The ‘trivial bound’ for  $\langle h^\sharp, h^\sharp \rangle$ , ignoring the cancellation derived above from the oscillation of the half-integral weight automorphy factor, is  $O(p^\theta)$ .

To illustrate the surprising power of (2.7), we now prove Theorem 5. Define  $V_0 : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$  by

$$V_0(u) := \int_{y=u}^\infty |W(y)|^2 \frac{dy}{y^2}.$$

By the asymptotic expansion of  $W$  near 0, there is  $u_0 \in (0, 1/e)$  so that  $V_0(u) \asymp \log(1/u)$  for  $|u| < u_0$ . Let  $C \geq 1$ . By (2.7), the final assertion of Section 3.4.4, Parseval, and the change of variables  $y \mapsto py/n$ , we obtain

$$\begin{aligned} 1 &\gg \|h^\sharp\|^2 \\ &\gg \frac{1}{Cp} \int_{x=0}^p \int_{y=u_0/C}^\infty |h^\sharp(x + iy)|^2 \frac{dx dy}{y^2} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{C} \sum_n \frac{|b(pn)|^2}{|n|} \int_{y=u_0/C}^\infty \left| W\left(\frac{ny}{p}\right) \right|^2 \frac{dy}{y^2} \\
 &= \frac{1}{Cp} \sum_n |b(pn)|^2 V_0\left(u_0 \frac{n}{Cp}\right) \\
 &\gg \frac{1}{Cp} \sum_{n:|n| \leq Cp} |b(pn)|^2 \log\left(u_0^{-1} \frac{Cp}{n}\right) \\
 &\geq \frac{1}{Cp} \sum_{n:|n| \leq Cp} |b(pn)|^2 \left(1 + \log\left(\frac{Cp}{n}\right)\right).
 \end{aligned}$$

We conclude by the  $L$ -value formula (2.3) for  $|b(d)|^2$ .

### 5. A triple product identity

In this section, we prove (2.10) after developing some preliminaries.

#### 5.1. Linear algebra lemma

In this subsection, we employ the temporary notation  $G := \text{SL}_2(\mathbb{F}_p)$ ,  $M = M_2(\mathbb{F}_p)$ . The group  $G$  acts on the  $\mathbb{F}_p$ -vector space  $M$  by conjugation. Let  $L := \left\{ \begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix} \right\}$  denote the strictly upper-triangular subspace of  $M$ . Let  $B$  denote the upper-triangular subgroup of  $G$ . Intrinsicly,  $B$  is the normalizer in  $G$  of  $L$ . We note, for future reference, that  $G$ ,  $B$ , and  $L$  are the images under reduction modulo  $p$  of  $\text{SL}_2(\mathbb{Z})$ ,  $\Gamma_0(p)$ , and  $pS(1/p)$ , respectively. We require the following calculation:

**Lemma 3.** *For  $x \in M$ , we have*

$$\#\{g \in G/B : x \in gLg^{-1}\} = \begin{cases} p + 1 & \text{if } x = 0, \\ 1 & \text{if } x \neq 0, \text{tr}(x) = \det(x) = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

*Proof.* Let  $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  denote the standard basis elements for the two-dimensional  $\mathbb{F}_p$ -vector space  $\mathbb{F}_p^2$ . The subspace  $L$  of  $M$  consists of those nilpotent elements  $x \in M$  that annihilate the line  $\ell := \mathbb{F}_p e_1$  spanned by  $e_1$ . The conjugate subspace  $gLg^{-1}$  thus consists of the nilpotent elements of  $M$  that annihilate the translated line  $g\ell$ . As  $g$  traverses  $G/B$ , the translated line  $g\ell$  traverses the set of all lines in  $\mathbb{F}_p^2$ . Thus the LHS of (5.1) vanishes unless  $x$  is nilpotent, in which case it is the number of lines that  $x$  annihilates.

In verifying (5.1), we may and shall assume that  $x$  is nilpotent, since otherwise both sides vanish. The claimed identity is then an immediate consequence of the following observations:

- The zero element annihilates all  $p + 1$  lines in  $\mathbb{F}_p^2$ .
- Any nilpotent element has a nontrivial kernel and hence annihilates some line.
- No nonzero nilpotent element annihilates two distinct lines.

□



### 5.2. Pushforward of the theta kernel

For  $w, z \in \mathbb{H}$ , set

$$\theta^\sharp(w, z) := \sum_{\gamma \in \Gamma_0(p) \backslash \mathrm{SL}_2(\mathbb{Z})} \theta(S(1/p), \gamma w, \gamma z).$$

**Lemma 4.** We have  $\theta^\sharp(w, z) = p\theta(z) \left( \frac{1}{p^{1/2}} \sum_{j \in \mathbb{Z}/p\mathbb{Z}} \theta(w, \frac{z+pj}{p^2}) + \theta(w, z) \right)$ .

*Proof.* Abbreviate  $\phi := \phi_{w,w,z}, \phi^0 := \phi_{w,z}^0$ . By definition,

$$\theta^\sharp(w, z) = \sum_{\gamma \in \Gamma_0(p) \backslash \mathrm{SL}_2(\mathbb{Z})} \sum_{\alpha \in \gamma^{-1}S(1/p)\gamma} \phi(\alpha), \quad \theta(w, z) = \sum_{\beta \in S^0} \phi^0(\beta).$$

For each  $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ , we have  $S \subseteq \gamma^{-1}S(1/p)\gamma \subseteq p^{-1}S$ . Conversely, let  $\alpha \in p^{-1}S$ . By applying the lemma of Section 5.1 to the image of  $p\alpha$  under reduction modulo  $p$ , we see that  $\alpha$  belongs to  $\gamma^{-1}S(1/p)\gamma$  for some  $\gamma \in \mathrm{SL}_2(\mathbb{Z})$  if and only if  $\mathrm{trace}(\alpha) \in \mathbb{Z}$  and  $\mathrm{det}(\alpha) \in p^{-1}\mathbb{Z}$ , and in that case,

$$\#\{\gamma \in \Gamma_0(p) \backslash \mathrm{SL}_2(\mathbb{Z}) : \alpha \in \gamma^{-1}S(1/p)\gamma\} = \begin{cases} 1 & \text{if } \alpha \notin S, \\ p+1 & \text{if } \alpha \in S. \end{cases}$$

Thus

$$\theta^\sharp(w, z) = \sum_{\substack{\alpha \in p^{-1}S: \\ \mathrm{tr}(\alpha) \in \mathbb{Z}, \\ \mathrm{det}(\alpha) \in p^{-1}\mathbb{Z}}} \phi(\alpha) + p \sum_{\alpha \in S} \phi(\alpha).$$

Using the decomposition (3.1) and the identity (3.2), we obtain

$$\begin{aligned} \sum_{\alpha \in S} \phi(\alpha) &= \sum_{m \in \mathbb{Z}, \beta \in S^0} y^{1/4} e(m^2 z) \phi^0(\beta) = \theta(z) \theta(w, z), \\ \sum_{\substack{\alpha \in p^{-1}S: \\ \mathrm{tr}(\alpha) \in \mathbb{Z}, \\ \mathrm{det}(\alpha) \in p^{-1}\mathbb{Z}}} \phi(\alpha) &= \theta(z) \sum_{\substack{\beta \in p^{-1}S^0: \\ \mathrm{det}(\beta) \in p^{-1}\mathbb{Z}}} \phi^0(\beta), \\ \sum_{\substack{\beta \in p^{-1}S^0: \\ \mathrm{det}(\beta) \in p^{-1}\mathbb{Z}}} \phi^0(\beta) &= \sum_{\substack{\beta \in S^0: \\ \mathrm{det}(\beta) \in p\mathbb{Z}}} \phi^0(p^{-1}\beta) = p^{1/2} \sum_{j \in \mathbb{Z}/p\mathbb{Z}} \theta(w, \frac{z+pj}{p^2}). \end{aligned}$$

These identities combine to yield the required identity. □

### 5.3. Period identities

We now prove (2.10). Note first that we may explicitly pushforward the  $L^2$ -mass of  $\varphi$  down to  $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$  before we integrate it against  $\Psi$ :

$$\langle \Psi\varphi, \varphi \rangle = \frac{1}{p+1} \int_{w \in \mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} \Psi(w) \sum_{\gamma \in \Gamma_0(p) \backslash \mathrm{SL}_2(\mathbb{Z})} |\varphi|^2(\gamma w) d\mu(w). \tag{5.2}$$

This identity motivates finding a formula for the inner sum over  $\gamma$ . To that end, we begin by applying the substitution  $z \mapsto -1/z$  to the integral representation for  $\overline{\varphi}(w_1)\varphi(w_2)$  given by the lemma of Section 3.5.

Using the formulas (2.8) and (3.3) describing the behavior of the integrand under  $z \mapsto -1/z$ , we obtain

$$\overline{\varphi(w_1)}\varphi(w_2) = \frac{\pm 1}{p} \int_{z \in \Gamma_0(1/p) \backslash \mathbb{H}} \theta(R(1/p); w_1, w_2, z) \varphi\left(\frac{z}{p}\right) d\mu(z). \tag{5.3}$$

Next, we observe that  $R(1/p) = \{\alpha/2 : \alpha \in S(1/p), \text{nr}(\alpha) \equiv 0 \pmod{4}\}$ . From this, the identity

$$\phi_{w_1, w_2, (z+pj)/4}(\alpha) = (1/4)e^{(pj \det(\alpha)/4)} \phi_{w_1, w_2, z}(\alpha/2),$$

and finite Fourier inversion, we deduce that

$$\theta(R(1/p); w_1, w_2, z) = \sum_{j \in \mathbb{Z}/4\mathbb{Z}} \theta(S(1/p); w_1, w_2, \frac{z+pj}{4}). \tag{5.4}$$

We choose a congruence subgroup  $\Gamma$  of  $\Gamma_0(1/p)$  that contains  $\{\pm 1\}$  but is otherwise small enough that each integral displayed below is well-defined. Since  $\varphi\left(\frac{z-pj}{p}\right) = \varphi\left(\frac{z}{p}\right)$ , the substitution  $z \mapsto z - pj$  followed by  $z \mapsto 4z$  gives

$$\begin{aligned} & \int_{z \in \Gamma \backslash \mathbb{H}} \theta(S(1/p); w_1, w_2, \frac{z+pj}{4}) \varphi\left(\frac{z}{p}\right) d\mu(z) \\ &= \int_{z \in \Gamma \backslash \mathbb{H}} \theta(S(1/p); w_1, w_2, \frac{z}{4}) \varphi\left(\frac{z}{p}\right) d\mu(z) \\ &= \int_{z \in \Gamma \backslash \mathbb{H}} \theta(S(1/p); w_1, w_2, z) \varphi\left(\frac{4z}{p}\right) d\mu(z) \end{aligned}$$

We introduce the shorthand  $\mathbb{E}_{z \in \Gamma \backslash \mathbb{H}}(\dots)$  for  $[\text{SL}_2(\mathbb{Z}) : \Gamma]^{-1} \int_{z \in \Gamma \backslash \mathbb{H}}(\dots)$ . Using the above computations, the identity  $[\text{SL}_2(\mathbb{Z}) : \Gamma_0(1/p)] = p + 1$ , and the formula (5.3), we deduce that

$$\overline{\varphi(w_1)}\varphi(w_2) = \frac{\pm 4(p+1)}{p} \mathbb{E}_{z \in \Gamma \backslash \mathbb{H}} \theta(S(1/p); w_1, w_2, z) \varphi\left(\frac{4z}{p}\right) d\mu(z). \tag{5.5}$$

Setting  $w_1 = w_2 =: w$  in (5.5) gives

$$\sum_{\gamma \in \Gamma_0(p) \backslash \text{SL}_2(\mathbb{Z})} |\varphi|^2(\gamma w) = \frac{\pm 4(p+1)}{p} \mathbb{E}_{z \in \Gamma \backslash \mathbb{H}} \theta^\sharp(w, z) \varphi\left(\frac{4z}{p}\right) d\mu(z).$$

Applying the lemma of Section 5.2, we obtain

$$\sum_{\gamma \in \Gamma_0(p) \backslash \text{SL}_2(\mathbb{Z})} |\varphi|^2(\gamma w) = \pm 4(p+1) \mathbb{E}_{z \in \Gamma \backslash \mathbb{H}} \theta(z) \frac{1}{p^{1/2}} \sum_{j \in \mathbb{Z}/p\mathbb{Z}} \theta(w, \frac{z+pj}{p^2}) \varphi\left(\frac{4z}{p}\right) d\mu(z). \tag{5.6}$$

(We have used here that  $\theta(z)\theta(w, z)$  is old at  $p$  and  $\varphi$  is new at  $p$  to discard the contribution of the second term in the lemma of Section 5.2.) We note that

$$\int_{w \in \text{SL}_2(\mathbb{Z}) \backslash \mathbb{H}} \Psi(w) \frac{1}{p^{1/2}} \sum_{j \in \mathbb{Z}/p\mathbb{Z}} \theta(w, \frac{z+pj}{p^2}) d\mu(w) = \overline{h^\sharp(w)}. \tag{5.7}$$

Combining (5.2) and (5.7) with (5.6) integrated over  $w$  against  $\Psi(w)$  gives

$$\langle \Psi\varphi, \varphi \rangle = \pm 4 \mathbb{E}_{z \in \Gamma \backslash \mathbb{H}} \theta(z) \overline{h^\sharp(z)} \varphi\left(\frac{4z}{p}\right) d\mu(z) = \pm 4 \langle \theta(z) \varphi\left(\frac{4z}{p}\right), h^\sharp(z) \rangle,$$

as required.

**6. The basic inequality**

We now prove (2.13). Set  $\Gamma := \Gamma_0(4/p)$ . Define  $F : \Gamma \backslash \mathbb{H} \rightarrow \mathbb{R}_{\geq 0}$  by  $F(z) := \text{ht}(z)^{1/2} |h^\#(z)|^2$ . By (2.1) and the estimate  $[\text{SL}_2(\mathbb{Z}) : \Gamma] \asymp p$ , our task (2.13) reduces to showing for  $T \geq 1$  that

$$\int_{\Gamma \backslash \mathbb{H}} F := \int_{z \in \Gamma \backslash \mathbb{H}} F(z) d\mu(z) \ll pT^{1/2} + p^{1/2}R, \tag{6.1}$$

where  $R$  is as in Section 2.9 and the implied constant is uniform in  $p, T$ . Using the fundamental domain from Section 3.4.4, we may write  $\int_{\Gamma \backslash \mathbb{H}} F = I_0 + \sum_{\gamma \in \mathcal{C}} I(\gamma)$ , where

$$I_0 := \int_{z \in \Gamma \backslash \mathbb{H} : \text{ht}(z) \leq T} F(z) d\mu(z) = \int_{\substack{z \in \Gamma \backslash \mathbb{H} \\ \text{ht}(z) \leq T}} \text{ht}(z)^{1/2} |h^\#(z)|^2 d\mu(z),$$

$$I(\gamma) := \int_{y=T}^\infty \int_{x=0}^{w(\gamma)} F(\gamma(x+iy)) \frac{dx dy}{y^2}$$

$$= \int_{y=T}^\infty y^{1/2} \int_{x=0}^{w(\gamma)} |h^\#(\gamma(x+iy))|^2 \frac{dx dy}{y^2}.$$

The adequate estimate  $I_0 \ll pT^{1/2}$  follows as indicated in Section 2.9 from the estimate  $\|h^\#\|^2 = \|h\|^2 \ll 1$ .

It remains to estimate  $I(\gamma_i)$  for  $i = 1, \dots, 6$ . We start with the most important case  $i = 1$ . Substituting the formula (2.6) for  $h^\#$  and appealing to Parseval followed by the substitution  $y \mapsto py/n$ , we obtain

$$I(\gamma_1) = p \int_{y=T}^\infty y^{1/2} \sum_n \frac{|b(pn)|^2}{|n|} \left| W\left(\frac{n}{p}y\right) \right|^2 \frac{dx dy}{y^2} \tag{6.2}$$

$$= p^{1/2} \sum_n \frac{|b(pn)|^2}{|n|^{1/2}} V\left(\frac{n}{pT}\right), \quad V(u) := \int_{y=au}^\infty y^{1/2} |W(y)|^2 \frac{dy}{y^2}. \tag{6.3}$$

The estimate (3.4) for  $W$  implies that  $V(u) \ll \min(1, |u|^{-100})$ , which leads to the adequate estimate  $I(\gamma_1) \ll p^{1/2}R$ .

Using Section 3.4.5, we may similarly estimate  $I(\gamma_2), I(\gamma_3)$ .

Since  $w(\gamma_i) \leq 4 = O(1)$  for  $i = 4, 5, 6$ , the ‘trivial bound’  $I(\gamma_i) \ll w(\gamma_i) \|h^\#\|_\infty^2 \ll pw(\gamma_i)$  suffices for our purposes.

**7. A converse estimate**

We finally prove Theorem 6. The non-constant Fourier components of  $\theta$  decay rapidly near the cusp, so we may find some fixed  $y_0 \geq 1$  so that

$$\theta(z) \gg y^{1/4} \text{ if } y = \text{Im}(z) \geq y_0. \tag{7.1}$$

Assuming (1.7) or (1.10), and arguing as in Section 6, we derive the lower bound

$$p^{1+o(1)} \gg p \|\theta h^\#\|^2 \gg \int_{y=y_0}^\infty y^{1/2} \int_{x=0}^p |h^\#(\gamma(x+iy))|^2 \frac{dx dy}{y^2} \tag{7.2}$$

$$= p^{1/2} \sum_n \frac{|b(pn)|^2}{|n|^{1/2}} V(y_0 n/p), \tag{7.3}$$

with  $V$  as defined in (6.3). Assuming that  $L(\Psi, \frac{1}{2}) \neq 0$ , the form  $h$  is nonzero, and its Whittaker function  $W$  is not identically zero on any interval. In particular,  $V(u) \gg 1$  for  $u \leq y_0$ . The required estimate (1.15) follows.

**Acknowledgements.** We gratefully acknowledge the support of NSF grant OISE-1064866 and SNF grant SNF-137488 during the work leading to this paper. Most of this article was written while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester, supported by the National Science Foundation under Grant No. DMS-1440140. We would like to thank the anonymous referee for helpful feedback and corrections, and Valentin Blomer, Philippe Michel, Etienne Le Masson, Maksym Radziwiłł, K. Soundararajan, Raphael Steiner, and Matthew Young for helpful feedback and encouragement.

**Conflict of Interest:** None.

## References

- [1] Joseph Bernstein and Andre Reznikov, ‘Subconvexity bounds for triple  $L$ -functions and representation theory’, *Ann. of Math. (2)* **172**(3): 1679–1718, 2010.
- [2] András Biró, ‘A relation between triple products of weight 0 and weight  $\frac{1}{2}$  cusp forms’, *Israel J. Math.* **182**: 61–101, 2011.
- [3] Valentin Blomer, ‘Period integrals and Rankin-Selberg  $L$ -functions on  $\mathrm{GL}(n)$ ’, *Geom. Funct. Anal.* **22**(3): 608–620, 2012.
- [4] Y. Colin de Verdière, ‘Ergodicité et fonctions propres du laplacien’, In *Bony-Sjöstrand-Meyer seminar, 1984–1985*, pages Exp. No. 13, **8** (École Polytech., Palaiseau, 1985).
- [5] 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.
- [6] *NIST Digital Library of Mathematical Functions*, <http://dlmf.nist.gov/>, Release 1.0.26 of 2020-03-15, F. W. J. Olver, A. B. Olde Daalhuis D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.
- [7] W. Duke, Ö. Imamoglu, and Á. Tóth, ‘Geometric invariants for real quadratic fields’, *Ann. of Math. (2)* **184**(3): 949–990, 2016.
- [8] W. Duke and H. Iwaniec, ‘Bilinear forms in the Fourier coefficients of half-integral weight cusp forms and sums over primes’, *Math. Ann.* **286**(4): 783–802, 1990.
- [9] Brooke Feigon and David Whitehouse, ‘Exact averages of central values of triple product  $L$ -functions’, *Int. J. Number Theory* **6**(7): 1609–1624, 2010.
- [10] D. R. Heath-Brown, ‘A mean value estimate for real character sums’, *Acta Arith.* **72**(3): 235–275, 1995.
- [11] Jeffrey Hoffstein and Paul Lockhart, ‘Coefficients of Maass forms and the Siegel zero’, *Ann. of Math. (2)* **140**(1): 161–181, 1994, with an appendix by Dorian Goldfeld, Jeffrey Hoffstein, and Daniel Lieman.
- [12] Atsushi Ichino, ‘Pullbacks of Saito-Kurokawa lifts’, *Invent. Math.* **162**(3): 551–647, 2005.
- [13] H. Iwaniec and P. Michel, ‘The second moment of the symmetric square  $L$ -functions’, *Ann. Acad. Sci. Fenn. Math.* **26**(2): 465–482, 2001.
- [14] Henryk Iwaniec, *Topics in classical automorphic forms*, volume **17** of *Graduate Studies in Mathematics* (American Mathematical Society, Providence, RI, 1997).
- [15] Henryk Iwaniec, *Spectral methods of automorphic forms*, volume **53** of *Graduate Studies in Mathematics* (American Mathematical Society, Providence, RI, 2e, 2002).
- [16] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume **53** of *American Mathematical Society Colloquium Publications* (American Mathematical Society, Providence, RI, 2004).
- [17] Junehyuk Jung, ‘Quantitative quantum ergodicity and the nodal domains of Hecke-Maass cusp forms’, *Comm. Math. Phys.* **348**(2): 603–653, 2016.
- [18] Svetlana Katok and Peter Sarnak, ‘Heegner points, cycles and Maass forms’, *Israel J. Math.* **84**(1-2): 193–227, 1993.
- [19] Etienne Le Masson and Tuomas Sahlsten, ‘Quantum ergodicity and Benjamini-Schramm convergence of hyperbolic surfaces’, *Duke Math. J.* **166**(18): 3425–3460, 2017.
- [20] S. Lester and M. Radziwiłł, ‘Quantum Unique Ergodicity for Half-Integral Weight Forms’, *ArXiv e-prints*, June 2016.
- [21] Wenzhi Luo and Peter Sarnak, ‘Quantum ergodicity of eigenfunctions on  $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}^2$ ’, *Inst. Hautes Études Sci. Publ. Math.* **(81)**: 207–237, 1995.
- [22] Wenzhi Luo and Peter Sarnak, ‘Mass equidistribution for Hecke eigenforms’, *Comm. Pure Appl. Math.* **56**(7): 874–891, 2003, dedicated to the memory of Jürgen K. Moser.
- [23] Wenzhi Luo and Peter Sarnak, ‘Quantum variance for Hecke eigenforms’, *Ann. Sci. École Norm. Sup. (4)* **37**(5): 769–799, 2004.
- [24] Philippe Michel and Akshay Venkatesh, ‘The subconvexity problem for  $\mathrm{GL}_2$ ’, *Publ. Math. Inst. Hautes Études Sci.* **(111)**: 171–271, 2010.
- [25] Paul D. Nelson, ‘Equidistribution of cusp forms in the level aspect’, *Duke Math. J.* **160**(3): 467–501, 2011.
- [26] Paul D. Nelson, ‘Evaluating modular forms on Shimura curves’, *Math. Comp.* **84**(295): 2471–2503, 2015.

- [27] Paul D. Nelson, 'Quantum variance on quaternion algebras, I', preprint, 2016.
- [28] Paul D. Nelson, 'The spectral decomposition of  $|\theta|^2$ ', preprint, 2016.
- [29] Paul D. Nelson, 'Quantum variance on quaternion algebras, II', preprint, 2017.
- [30] Paul D. Nelson, 'Quantum variance on quaternion algebras, III', preprint, 2019.
- [31] Paul D. Nelson, 'Subconvex equidistribution of cusp forms: Reduction to Eisenstein observables', *Duke Math. J.* **168**(9): 1665–1722, 2019.
- [32] Yannan Qiu, 'Periods of Saito-Kurokawa representations', *Int. Math. Res. Not. IMRN* (24): 6698–6755, 2014.
- [33] Jean Raimbault, 'On the convergence of arithmetic orbifolds', *Ann. Inst. Fourier (Grenoble)* **67**(6): 2547–2596, 2017.
- [34] Peter Sarnak, 'Fourth moments of Größencharakteren zeta functions', *Comm. Pure Appl. Math.* **38**(2): 167–178, 1985.
- [35] A. I. Schnirelman, 'Ergodic properties of eigenfunctions', *Uspehi Mat. Nauk* **29**(6(180)): 181–182, 1974.
- [36] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*. Publications of the Mathematical Society of Japan, No. **11** (Iwanami Shoten, Publishers, Tokyo, 1971), Kanô Memorial Lectures, No. 1.
- [37] Goro Shimura, 'On modular forms of half integral weight', *Ann. of Math. (2)* **97**: 440–481, 1973.
- [38] Takuro Shintani, 'On construction of holomorphic cusp forms of half integral weight', *Nagoya Math. J.* **58**: 83–126, 1975.
- [39] K. Soundararajan and Matthew P. Young, 'The second moment of quadratic twists of modular  $L$ -functions', *J. Eur. Math. Soc. (JEMS)* **12**(5): 1097–1116, 2010.
- [40] Raphael S. Steiner, 'Sup-norm of Hecke-Laplace Eigenforms on  $S^3$ ', *arXiv e-prints* page arXiv:1811.03949, Nov 2018.
- [41] Nicolas Templier and Jacob Tsimerman, 'Non-split sums of coefficients of  $GL(2)$ -automorphic forms', *Israel J. Math.* **195**(2): 677–723, 2013.
- [42] Akshay Venkatesh, 'Sparse equidistribution problems, period bounds and subconvexity', *Ann. of Math. (2)* **172**(2): 989–1094, 2010.
- [43] Thomas C. Watson, 'Rankin triple products and quantum chaos', arXiv.org:0810.0425, 2008.
- [44] Steven Zelditch, 'Uniform distribution of eigenfunctions on compact hyperbolic surfaces', *Duke Math. J.* **55**(4): 919–941, 1987.
- [45] Peng Zhao, 'Quantum variance of Maass-Hecke cusp forms', *Comm. Math. Phys.* **297**(2): 475–514, 2010.