

# NUMERICAL COMPUTATIONS WITH THE TRACE FORMULA AND THE SELBERG EIGENVALUE CONJECTURE

ANDREW R. BOOKER AND ANDREAS STRÖMBERGSSON

ABSTRACT. We verify the Selberg eigenvalue conjecture for congruence groups of small squarefree conductor, improving on a result of Huxley [20]. The main tool is the Selberg trace formula which, unlike previous geometric methods, allows for treatment of cases where the eigenvalue  $\frac{1}{4}$  is present. We present a few other sample applications, including the classification of even 2-dimensional Galois representations of small squarefree conductor.

## 1. INTRODUCTION

Let  $\mathcal{H} := \{x + iy \mid x, y \in \mathbb{R}, y > 0\}$  be the upper half plane,  $\Delta := -y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)$  the hyperbolic Laplacian, and  $\Gamma \supset \Gamma(N) := \{\gamma \in \mathrm{SL}(2, \mathbb{Z}) \mid \gamma \equiv I \pmod{N}\}$  a congruence subgroup of  $\mathrm{SL}(2, \mathbb{Z})$  of level  $N$ . In [32], Selberg conjectured that  $\lambda_1(\Gamma)$ , the smallest eigenvalue of  $\Delta$  acting on the space of cusp forms  $L^2_{\mathrm{cusp}}(\Gamma \backslash \mathcal{H})$ , is bounded below by  $\frac{1}{4}$ . In the same paper, he proved the lower bound  $\lambda_1(\Gamma) \geq \frac{3}{16}$ ; this has steadily been improved over recent years to  $\lambda_1(\Gamma) \geq \frac{1}{4} - \left(\frac{7}{64}\right)^2 = 0.238037109375$  by making use of results toward Langlands' functoriality conjectures [22]. Despite this recent progress, the current record in terms of affirming Selberg's conjecture is a theorem of Huxley [20] from 1985 establishing the cases with  $N \leq 18$ .

In this paper, we attack the problem of verifying the conjecture by a method substantially different from that of Huxley. In particular, our method, which relies on the Selberg trace formula, allows us to treat cases where the lower bound is exact, i.e. where the eigenvalue  $\frac{1}{4}$  is present<sup>1</sup>; according to Artin's conjecture, most cases of which are known in the present context, this happens whenever there is an even 2-dimensional Galois representation of Artin conductor  $N$ . (Conversely, it is believed that all forms of eigenvalue  $\frac{1}{4}$  are associated to Galois representations; see [24, p. 2].) Our main result is the following (recall  $\Gamma_1(N) := \{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z}) \mid a \equiv d \equiv 1, c \equiv 0 \pmod{N}\}$ ):

**Theorem 1.** *Selberg's conjecture is true for  $\Gamma = \Gamma_1(N)$  for all squarefree  $N < 857$ .*

The statement of the result in terms of  $\Gamma_1(N)$  as opposed to  $\Gamma(N)$  essentially means that we consider forms of *conductor*  $N$  rather than *level*  $N$ ; this is more natural from several perspectives, including that of the trace formula. On the other hand, the forms of level  $N$  may be regarded as a subspace of the forms on  $\Gamma_1(N^2) \backslash \mathcal{H}$  (cf., e.g., [29, p. 114]), which we will treat in a future paper. In the present paper, we concentrate on the case of squarefree conductor. In Section 2 we present a detailed derivation of the trace formula for this case. While well understood from a more abstract point of view, to the best of our knowledge no

---

<sup>1</sup>By contrast, Huxley's geometric method, when applicable, always produces a strict lower bound. Upon learning of the existence of forms of eigenvalue  $\frac{1}{4}$ , Huxley writes, he "perceived the futility of numerical methods" [21, §4]. Thus, the inability to deal with the eigenvalue  $\frac{1}{4}$  forms may explain the lack of more recent results on the subject.

explicit version of this formula (suitable for numerical computation) has formerly appeared in print. One side benefit of our numerical method is that we are able to test the final formula for correctness.

As mentioned above, the forms of eigenvalue  $\frac{1}{4}$  are closely connected to 2-dimensional Galois representations. To establish Selberg’s conjecture, our method depends on finding such representations when they exist; this is carried out for dihedral representations and more general “CM” forms in Section 3, and for exotic (non-dihedral) representations in Section 5. Conversely, the trace formula can be used to bound the dimension of the  $\lambda = \frac{1}{4}$  eigenspace, and thus (assuming Artin’s conjecture) obtain information about Galois representations. In Section 5, we use this idea to classify the even 2-dimensional representations of small squarefree conductor; cf. Theorem 5. We return to this problem in greater depth in [9].

In Section 4, we develop a general method for optimizing the test functions to use in trace-like formulae. (See [6] for an application to the “explicit formula” in the context of  $L$ -functions.) The method allows us to extend significantly the computational range that we could achieve otherwise. Moreover, it yields a technique for estimating low eigenvalues. In particular, we furnish a small ingredient needed in [11] to certify the first few eigenvalues for  $\Gamma(1)$ ; namely, we prove that there are exactly ten such eigenvalues  $\leq 400$ , cf. Proposition 4.4. We consider this and related questions, as well as the study of the distribution of eigenvalues in the level aspect, in [10].

Finally, in Section 6 we discuss the complexity of the various applications of our method. In particular, while computing with the trace formula has exponential complexity in the level, we give a heuristic polynomial-time algorithm for verifying Selberg’s conjecture by combining the methods of this paper with those of [11].

#### ACKNOWLEDGEMENTS

The first author was supported by an NSF postdoctoral fellowship. Some of the initial work on this project was carried out at the Newton Institute in Cambridge; we thank them for their generous support. Also, we thank D. Doud, A. Jehanne, M. Krishnamurthy, P. Michel, P. Sarnak, F. Strömberg, and A. Venkatesh for helpful discussions. Finally, special thanks to Norm Hurt for his continued interest in our work.

## 2. THE SELBERG TRACE FORMULA

For our numerical study it will be important to desymmetrize the given space of functions as far as possible, in order to obtain as low eigenvalue density as possible in each individual application of the trace formula. Recall that the space  $L^2(\Gamma_1(N)\backslash\mathcal{H})$  decomposes as a direct sum of spaces of functions invariant under  $\Gamma_0(N)$  with nebentypus character  $\chi$ , where  $\chi$  runs through all even Dirichlet characters modulo  $N$  (cf., e.g., [29, Lemma 4.3.1]). Each such space can be further decomposed into spaces of odd and even functions, and newforms and oldforms.

Our goal in this section is to derive a maximally explicit and ready-to-apply version of the Selberg trace formula for the space of even (or odd) Maass newforms invariant under  $\Gamma_0(N)$  ( $N$  squarefree) with an arbitrary nebentypus character  $\chi$ . The main technical difficulty comes from our wish to separate the even and odd parts of the spectrum. For this we find it convenient to treat, as an intermediate step, the case of a general cofinite group of isometries (not necessarily orientation preserving) of the hyperbolic plane, with an arbitrary character. This is done in Section 2.1, see (2.37) and Theorem 2 for the resulting trace formula. This

formula generalizes one given by Venkov [42] for the case of Dirichlet or von Neumann boundary problems on certain (compact or non-compact) polygons in the hyperbolic plane. (Cf. Remark 2.6 below; in particular note that our case involves a considerably more delicate interplay with the continuous spectrum.) The general trace formula in (2.37) will be applied again in a sequel to this paper, [40], to the case of  $\Gamma_0(N)$  with  $N$  not squarefree. Formula (2.37) might also prove to be useful for numerical applications of the trace formula in other cases, e.g. noncongruence subgroups of  $\mathrm{PSL}(2, \mathbb{Z})$  (cf. [35, Ch. 3]; note the prominent role of mirror symmetries therein).

For previous work on the Selberg trace formula for  $\Gamma_0(N)$  with  $\chi = 1$  and without separation of the odd and even parts of the spectrum, see Hejhal [18, Ch. 11] and Huxley [19], where the cuspidal contribution to the trace formula is computed.

**2.1. The case of a general group of isometries.** We first fix some notation. The group of all isometries (orientation preserving or not) of  $\mathcal{H}$  can be identified with  $G = \mathrm{PGL}(2, \mathbb{R})$ , where the action is defined by

$$T(z) = \begin{cases} \frac{az+b}{cz+d} & \text{if } ad - bc > 0 \\ \frac{\bar{a}\bar{z}+b}{\bar{c}\bar{z}+d} & \text{if } ad - bc < 0, \end{cases} \quad \text{for } T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G.$$

The group of orientation preserving isometries,  $G^+ = \mathrm{PSL}(2, \mathbb{R})$ , is a subgroup of index 2 in  $G$ . We write  $G^- = G - G^+$  for the other coset in  $G$ . One easily shows that any element  $T \in G$ ,  $T \neq I$  is conjugate in  $G$  to a unique element of the form (with  $\gamma > 1$  or  $\gamma < -1$ , and  $0 < \theta \leq \frac{\pi}{2}$ ):

$$(2.1) \quad \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{k}(\theta) := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Geometrically, in the first case  $T$  is a hyperbolic map (if  $\gamma > 1$ ) or a glide reflection (if  $\gamma < -1$ ); in the second case  $T$  is a reflection, in the third case an elliptic map, and in the fourth case a parabolic map.

Now let  $\Gamma$  be a discrete subgroup of  $G$  such that the surface  $\Gamma \backslash \mathcal{H}$  has finite area, and let  $\chi$  be a (unitary) character on  $\Gamma$ . We set  $\Gamma^+ := \Gamma \cap G^+$  and assume  $\Gamma^+ \neq \Gamma$ . (Otherwise the desired trace formula is already stated in [18, Ch. 6].) We let  $L^2(\Gamma \backslash \mathcal{H}, \chi)$  be the Hilbert space of functions  $f : \mathcal{H} \rightarrow \mathbb{C}$  satisfying the automorphy relation  $f(Tz) \equiv \chi(T)f(z)$ ,  $\forall T \in \Gamma$ , and  $\int_{\Gamma \backslash \mathcal{H}} |f|^2 d\mu < \infty$ . The trace formula concerns the spectrum of the Laplace operator  $\Delta = -y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$  on  $L^2(\Gamma \backslash \mathcal{H}, \chi)$ . We let  $\phi_1, \phi_2, \dots$  be any orthonormal basis of the discrete spectrum of  $\Delta$ , i.e.  $\phi_n \in C^\infty(\mathcal{H}) \cap L^2(\Gamma, \chi)$  and  $\Delta \phi_n = \lambda_n \phi_n$ , say with increasing eigenvalues  $0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ . We also let  $r_n = \sqrt{\lambda_n - \frac{1}{4}} \in \mathbb{R}^+ \cup i[-\frac{1}{2}, 0]$ . By abuse of terminology, we may sometimes refer to  $r_n$  as the eigenvalue.

Note that  $\Gamma^+$  is a cofinite Fuchsian group, and for this group we will use similar notation to [18, p. 268]: Let  $\mathcal{F} \subset \mathcal{H}$  be a canonical (closed) fundamental domain for  $\Gamma^+ \backslash \mathcal{H}$ , and let  $\eta_1, \dots, \eta_\kappa$  (where  $\kappa \geq 0$ ) be the vertices of  $\mathcal{F}$  along  $\partial \mathcal{H} = \mathbb{R} \cup \{\infty\}$ . Since  $\mathcal{F}$  is canonical,  $\eta_1, \dots, \eta_\kappa$  are  $\Gamma$ -inequivalent. For each  $k \in \{1, \dots, \kappa\}$  we choose  $N_k \in G^+$  such that  $N_k(\eta_k) = \infty$  and such that the stabilizer  $\Gamma_{\eta_k}^+$  is  $[T_k]$ , where  $T_k := N_k^{-1}S^{-1}N_k$ , with  $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ . By modifying  $N_k$  we can also ensure that

$$(2.2) \quad N_k(\mathcal{F}) \cap \{z \in \mathcal{H} \mid \mathrm{Im} z \geq Y\} = [0, 1] \times [Y, \infty)$$

holds for all  $k \in \{1, \dots, \kappa\}$  and all  $Y \geq B_0$ , where  $B_0 = B_0(\Gamma^+) > 1$  is a constant fixed once and for all. (The final trace formula (2.37) is valid regardless of the assumption (2.2), since all data appearing in (2.37) are invariant under any replacement  $N_k \mapsto \begin{pmatrix} 1 & u_k \\ 0 & 1 \end{pmatrix} N_k$  with  $u_k \in \mathbb{R}$ ; however, (2.2) will be convenient in our proof of (2.37).) For  $Y \geq B_0$ , the corresponding cuspidal region in  $\mathcal{F}$  is labeled  $\mathcal{C}_{kY}$ :

$$(2.3) \quad \mathcal{C}_{kY} = N_k^{-1}([0, 1] \times [Y, \infty)) \subset \mathcal{F}.$$

We then define

$$(2.4) \quad \mathcal{F}_Y = \mathcal{F} - \bigcup_{k=1}^{\kappa} \mathcal{C}_{kY}.$$

This is a bounded region. Also  $\mathcal{C}_{kY} \cap \mathcal{C}_{jY} = \emptyset$  for all  $j \neq k$ .

Let us fix, once and for all, an element  $V \in \Gamma - \Gamma^+$ . We introduce some notation in order to describe in a precise way the action of  $V$  on the cusps: For each  $j \in \{1, \dots, \kappa\}$ ,  $VT_jV^{-1}$  is a parabolic element in  $\Gamma^+$ , and hence there are  $k(j) \in \{1, \dots, \kappa\}$ ,  $U_j \in \Gamma^+$  and  $n(j) \in \mathbb{Z} - \{0\}$  such that  $VT_jV^{-1} = U_j T_{k(j)}^{n(j)} U_j^{-1}$ . (Here  $k(j)$  and  $n(j)$  are uniquely determined, and  $U_j$  is determined up to right shifts with  $T_{k(j)}$ ; we *fix* one choice of  $U_j$  from now on.) This implies that  $g = N_{k(j)} U_j^{-1} V N_j^{-1} \in G^-$  satisfies  $gS^{-1}g^{-1} = S^{-n(j)}$ , and one shows by a quick computation that this implies  $n(j) < 0$  and

$$N_{k(j)} U_j^{-1} V N_j^{-1} = \begin{pmatrix} n(j) & * \\ 0 & 1 \end{pmatrix} \quad \text{in } G.$$

A similar identity holds with  $j$  replaced by  $k(j)$ , and multiplying these two we obtain:

$$U_{k(j)}^{-1} V U_j^{-1} V = N_{k(k(j))}^{-1} \begin{pmatrix} n(j)n(k(j)) & * \\ 0 & 1 \end{pmatrix} N_j.$$

It follows that  $U_{k(j)}^{-1} V U_j^{-1} V (\eta_j) = \eta_{k(k(j))}$ , and since  $U_{k(j)}^{-1} V U_j^{-1} V \in \Gamma \cap G^+ = \Gamma^+$  this forces  $k(k(j)) = j$  and  $U_{k(j)}^{-1} V U_j^{-1} V$  to be parabolic. Hence  $n(j)n(k(j)) = 1$ . Thus we have proved:

$$(2.5) \quad \forall j \in \{1, \dots, \kappa\} : \quad k(k(j)) = j, \quad n(j) = -1, \quad V = U_j N_{k(j)}^{-1} \begin{pmatrix} -1 & x_j \\ 0 & 1 \end{pmatrix} N_j,$$

for some  $x_j \in \mathbb{R}$ . Again studying the product  $\begin{pmatrix} -1 & x_{k(j)} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & x_j \\ 0 & 1 \end{pmatrix}$  we see that

$$(2.6) \quad \forall j \in \{1, \dots, \kappa\} : \quad U_{k(j)}^{-1} V U_j^{-1} V = T_j^{x_{k(j)} - x_j}, \quad \text{where necessarily } x_{k(j)} - x_j \in \mathbb{Z}.$$

Let  $C_{\Gamma, \chi} = \{j \in \{1, \dots, \kappa\} \mid \chi(T_j) = 1\}$  be the set of indices of *open* cusps. Note that  $j \in C_{\Gamma, \chi}$  implies  $k(j) \in C_{\Gamma, \chi}$ , since  $T_{k(j)} = (U_j^{-1} V) T_j^{-1} (U_j^{-1} V)^{-1}$ .

For  $j \in C_{\Gamma, \chi}$ , let  $E_j(z, s, \chi)$  be the Eisenstein series for  $\langle \Gamma^+, \chi \rangle$  associated to the cusp  $\eta_j$ , viz.

$$(2.7) \quad E_j(z, s, \chi) = \sum_{W \in [T_j] \backslash \Gamma^+} \chi(W^{-1}) (\text{Im } N_j W z)^s \quad (\text{Re } s > 1),$$

continued meromorphically to all  $s \in \mathbb{C}$ . Using (2.5) (for  $k(j)$  in place of  $j$ ) one proves, directly from the definition (2.7),

$$(2.8) \quad E_j(Vz, s, \chi) = \chi(U_{k(j)}) \cdot E_{k(j)}(z, s, \chi).$$

We define a natural  $\Gamma$ -analog of the Eisenstein series:

$$(2.9) \quad E_j^\Gamma(z, s, \chi) := \sum_{W \in [T_j] \backslash \Gamma} \chi(W^{-1}) (\text{Im } N_j W z)^s \quad (\text{Re } s > 1).$$

Clearly this function is  $(\Gamma, \chi)$ -automorphic, i.e.  $E_j^\Gamma(Tz, s, \chi) = \chi(T)E_j^\Gamma(z, s, \chi)$  for all  $T \in \Gamma$ . Using  $\Gamma = \Gamma^+ \sqcup \Gamma^+V$  and (2.8) we find that

$$(2.10) \quad E_j^\Gamma(z, s, \chi) = E_j(z, s, \chi) + \chi(V^{-1}U_{k(j)})E_{k(j)}(z, s, \chi).$$

Hence  $E_j^\Gamma(z, s, \chi)$  has a meromorphic continuation to all  $s \in \mathbb{C}$ .

We are now ready to start the development of the trace formula. We fix  $\delta > 0$ , and take  $h(r)$  to be an arbitrary even analytic function in the strip  $|\operatorname{Im} r| \leq \frac{1}{2} + \delta$  such that  $h(r) \in \mathbb{R}$  for  $r \in \mathbb{R}$  and  $h(r) = O(e^{-15|r|})$  throughout the strip. We then form the usual transforms (cf., e.g., [17, Ch. 1 (Prop 4.1)] or [18, pp. 168–169]):

$$(2.11) \quad g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \cdot e^{-iru} dr$$

$$(2.12) \quad Q(x) = g(u) \quad \text{for } x \geq 0, \quad x = e^u + e^{-u} - 2, \quad u \in \mathbb{R}$$

$$(2.13) \quad \Phi(t) = -\frac{1}{\pi} \int_t^\infty \frac{Q'(x)}{\sqrt{x-t}} dx = -\frac{2}{\pi} \int_0^\infty Q'(t+u^2) du, \quad t \geq 0$$

$$(2.14) \quad k(z, w) = \Phi \left[ \frac{|z-w|^2}{\operatorname{Im} z \cdot \operatorname{Im} w} \right], \quad \text{for } z, w \in \mathcal{H}.$$

The pretrace formula in our case reads as follows:

$$(2.15) \quad K(z, w, \chi) := \sum_{T \in \Gamma} \chi(T)k(z, Tw) \\ = \sum_{n \geq 1} h(r_n) \phi_n(z) \overline{\phi_n(w)} + \frac{1}{8\pi} \sum_{j \in C_{\Gamma, \chi}} \int_{-\infty}^{\infty} h(t) E_j^\Gamma(z, \frac{1}{2} + it, \chi) \overline{E_j^\Gamma(w, \frac{1}{2} + it, \chi)} dt.$$

Here all sums and integrals are uniformly absolutely convergent for  $(z, w)$  in  $\mathcal{H} \times \mathcal{H}$  compacta.

Formula (2.15) can be proved by generalizing the usual development for cofinite Fuchsian groups (cf. e.g. [18, Ch. 6,8]) to the slightly more general case of  $\Gamma \subset G$ , but it is more quickly derived as a *consequence* of [18, Ch. 6,8]: Define  $I_V : L^2(\Gamma^+, \chi) \rightarrow L^2(\Gamma^+, \chi)$  by  $I_V f(z) := \chi(V)^{-1} f(Vz)$ . Then  $I_V$  is a self-adjoint operator with  $I_V^2 = \operatorname{Id}$  and it commutes with the Laplace operator  $\Delta$ . Hence we may fix an orthonormal basis of discrete eigenfunctions  $\tilde{\phi}_1, \tilde{\phi}_2, \dots$  of  $\Delta$  in  $L^2(\Gamma^+, \chi)$  such that  $I_V \tilde{\phi}_n = s_n \tilde{\phi}_n$  ( $s_n \in \{1, -1\}$ ) for all  $n$ . Clearly then our basis  $\phi_1, \phi_2, \dots$  of  $L^2(\Gamma, \chi)$  can be chosen as  $\{\sqrt{2} \cdot \tilde{\phi}_n \mid s_n = 1\}$ . Now the pretrace formula for  $L^2(\Gamma^+, \chi)$  (see [18, p. 303 (4.1)], as well as [18, pp. 167 (13.2), 188-9 (Prop 13.6(g)), 306 (item 8)]) gives the spectral expansion of  $K^+(z, w, \chi) := \sum_{T \in \Gamma^+} \chi(T)k(z, Tw)$  in terms of  $\tilde{\phi}_n$  and  $E_j(\cdot, \frac{1}{2} + it, \chi)$ ; (2.15) is obtained directly from this formula upon using  $K(z, w, \chi) = K^+(z, w, \chi) + \chi(V)K^+(z, Vw, \chi)$  and (2.8), (2.10). In the argument just given the absolute convergence on compacta on the right hand side of (2.15) follows from [18, pp. 100, 188 (Prop 13.6(d)), 301(a,d), 315(ii)]. Regarding the convergence of  $\sum_{T \in \Gamma} \chi(T)k(z, Tw)$  we have:

**Lemma 2.1.** *Fix  $\delta > 0$ . If  $\Phi(t)$  is any continuous real-valued function on  $[0, \infty)$  such that  $|\Phi(t)| \leq A(t+4)^{-\delta}$  for all  $t \geq 0$ , and  $k(z, w) = \Phi(\frac{|z-w|^2}{\operatorname{Im} z \cdot \operatorname{Im} w})$ , then the sum  $K(z, w, \chi) =$*

$\sum_{T \in \Gamma} \chi(T)k(z, Tw)$  is uniformly absolutely convergent for  $(z, w)$  in  $\mathcal{H} \times \mathcal{H}$  compacta. Furthermore, for each  $j \in \{1, \dots, \kappa\}$ , writing  $F_j = \{T \in \Gamma \mid T(\eta_j) = \eta_j\}$ , we have:

$$\sum_{T \in \Gamma - F_j} |k(z, Tz)| = O((\operatorname{Im} N_j z)^{-2\delta}) \quad \text{as } \operatorname{Im} N_j z \rightarrow \infty.$$

Note that the assumption  $|\Phi(t)| \leq A(t+4)^{-\delta}$  certainly holds for  $\Phi(t)$  as in (2.13); cf. [18, pp. 168-9].

*Proof.* Since  $\Gamma = \Gamma^+ \cup \Gamma^+V$ , the first claim follows from the corresponding statement for  $K^+(z, w, \chi)$ , and this is derived from [18, p. 27 (Prop 5.1)], cf. [18, p. 31 (Prop 6.2)]. The second claim is proved by using  $\Gamma = \Gamma^+ \cup \Gamma^+V$ , (2.5), and the summation technique in [18, pp. 92-93]. (The details are exactly as in [37, Lemma 3.4, Lemma 3.5].)  $\square$

Let us write

$$(2.16) \quad H(z, w, \chi) = \frac{1}{8\pi} \sum_{j \in C_{\Gamma, \chi}} \int_{-\infty}^{\infty} h(t) E_j^\Gamma(z, \frac{1}{2} + it, \chi) \overline{E_j^\Gamma(w, \frac{1}{2} + it, \chi)} dt;$$

$$(2.17) \quad K_0(z, w, \chi) = K(z, w, \chi) - H(z, w, \chi).$$

Then by (2.15),  $K_0(z, w, \chi) = \sum_{n \geq 1} h(r_n) \phi_n(z) \overline{\phi_n(w)}$ . We now let  $w = z$  and integrate the last identity over  $\Gamma \setminus \mathcal{H}$ ; changing the order of integration and summation (as is justified by [18, pp. 188 (Prop 13.6(c)), 306 (item 8)]), we obtain

$$\sum_{n \geq 1} h(r_n) = \int_{\Gamma \setminus \mathcal{H}} K_0(z, z, \chi) d\mu(z).$$

Substituting here (2.17) and  $K(z, w, \chi) = \sum_{T \in \Gamma} \chi(T)k(z, Tw)$ , we wish to try to change the order of summation. We split the sum into a cuspidal and non-cuspidal part: Define

$$(2.18) \quad M = \{T \in \Gamma \mid T \text{ has no cusp [of } \Gamma^+ \text{] as a fixpoint}\}; \quad M_I = M \cup \{I\}.$$

Then since  $M \subset \Gamma - \cup_{j=1}^{\kappa} F_j$  it follows from (2.15), (2.17) and Lemma 2.1 that

$$(2.19) \quad \sum_{n \geq 1} h(r_n) = \sum_{T \in M_I} \int_{\Gamma \setminus \mathcal{H}} \chi(T)k(z, Tz) d\mu(z) \\ + \int_{\Gamma \setminus \mathcal{H}} \left( \sum_{T \in \Gamma - M_I} \chi(T)k(z, Tz) - H(z, z, \chi) \right) d\mu(z)$$

where all sums and integrals are absolutely convergent. (Regarding the last term, this should be interpreted to mean  $\int_{\Gamma \setminus \mathcal{H}} |\sum_{T \in \Gamma - M_I} \chi(T)k(z, Tz) - H(z, z, \chi)| d\mu(z) < \infty$ ; in fact by Lemma 2.1 and [18, pp. 188 (Prop 13.6(c)), 306 (item 8)] there is a constant  $0 < \beta < \frac{1}{2}$  such that

$$(2.20) \quad \left| \sum_{T \in \Gamma - M_I} \chi(T)k(z, Tz) - H(z, z, \chi) \right| = O((\operatorname{Im} N_j z)^{1-\beta}) \quad \text{as } \operatorname{Im} N_j z \rightarrow \infty,$$

for any  $j \in \{1, \dots, \kappa\}$ .)

Now the first sum on the right hand side of (2.19) is easily treated. Note that by definition  $M$  does not contain the identity or any parabolic elements. Recall the normal forms in (2.1); for  $T \in G$  hyperbolic, a glide reflection or a reflection we define  $N(T) := |\gamma|$  where  $\gamma$  is the unique real number with  $\gamma > 1$  or  $\gamma \leq -1$  such that  $T \sim \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}$  (thus  $N(T) = 1$  iff.  $T$  is a

reflection); if  $T$  is elliptic we let  $\theta(T) := \theta$  be the unique number in  $(0, \frac{\pi}{2}]$  such that  $T \sim \mathbf{k}(\theta)$ . We also let  $Z_\Gamma(T)$  denote the centralizer of  $T$  in  $\Gamma$ .

**Proposition 2.2.** *For each non-elliptic  $T \in M$  there exists some hyperbolic element or glide reflection  $T_0 \in Z_\Gamma(T)$  such that the infinite cyclic group  $[T_0] = \{T_0^n \mid n \in \mathbb{Z}\}$  has index 1, 2 or 4 in  $Z_\Gamma(T)$ . We now have, for each admissible  $h(r)$ :*

$$(2.21) \quad \sum_{T \in M_I} \int_{\Gamma \setminus \mathcal{H}} \chi(T) k(z, Tz) d\mu(z) = \frac{\mu(\Gamma \setminus \mathcal{H})}{4\pi} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr$$

$$+ \sum_{\substack{\{T\} \subset M, \\ \text{non-ell}}} \frac{\log N(T_0)}{[Z_\Gamma(T) : [T_0]]} \cdot \frac{\chi(T) \cdot g(\log N(T))}{N(T)^{1/2} - \text{sgn}(\det T) \cdot N(T)^{-1/2}}$$

$$+ \sum_{\substack{\{T\} \subset M, \\ \text{elliptic}}} \frac{\chi(T)}{2|Z_\Gamma(T)| \cdot \sin \theta(T)} \cdot \int_{-\infty}^{\infty} \frac{e^{-2\theta(T)r}}{1 + e^{-2\pi r}} \cdot h(r) dr.$$

The  $\{T\}$  sums are taken over distinct  $\Gamma$ -conjugacy classes; the first sum is taken over all non-elliptic conjugacy classes, i.e.  $T$  hyperbolic or a reflection or a glide reflection. The various sums and integrals are all convergent with good majorants. The elliptic sum is finite.

We remark that the factor  $\frac{\log N(T_0)}{[Z_\Gamma(T) : [T_0]]}$  in line 2 of (2.21) is independent of the choice of  $T_0 \in Z_\Gamma(T)$ ; indeed, one easily verifies that this ratio is well defined and takes the same value for any  $T_0 \in Z_\Gamma(T)$  which is hyperbolic or a glide reflection.

*Proof.* It follows from  $\Gamma = \Gamma^+ \cup \Gamma^+ V$  and (2.5) that  $\Gamma$  permutes the set of  $(\Gamma^+)$ -cusps; hence the subset  $M \subset \Gamma$  is seen to consist of full  $\Gamma$ -conjugacy classes. Collecting these conjugacy classes we obtain (cf., e.g., [17, Ch. 1.5])

$$(2.22) \quad \sum_{T \in M \cup \{I\}} \int_{\Gamma \setminus \mathcal{H}} \chi(T) k(z, Tz) d\mu(z) = \sum_{\{T\} \subset M \cup \{I\}} \chi(T) \int_{Z_\Gamma(T) \setminus \mathcal{H}} k(z, Tz) d\mu(z).$$

As in [17, pp. 27–28] we see that the contribution from  $T = I$  is given by the first term on the right hand side of (2.21).

Now fix a conjugacy class  $\{T\} \subset M$ ; we first assume that  $T$  is a reflection. Then after an auxiliary conjugation we may assume  $T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ . The centralizer of  $T$  in  $G$  is  $Z_G(T) = \{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \mid a \in \mathbb{R}^\times \} \cup \{ \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix} \mid b \in \mathbb{R}^\times \}$ . We claim that  $Z_\Gamma(T)$  must contain some hyperbolic element. The square of a glide reflection is hyperbolic, and the product  $\begin{pmatrix} 0 & b_1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & b_2 \\ 1 & 0 \end{pmatrix}$  is hyperbolic or a glide reflection whenever  $|b_1| \neq |b_2|$ ; hence if  $Z_\Gamma(T)$  did *not* contain any hyperbolic elements then we would have either  $Z_\Gamma(T) = \{I, T\}$  or  $Z_\Gamma(T) = \{I, T, \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -b \\ 1 & 0 \end{pmatrix}\}$  for some fixed  $b > 0$ . Then there would exist a fundamental region for  $Z_\Gamma(T) \setminus \mathcal{H}$  containing all of the domain  $D = \{z = x + iy \mid x > 0, y > \sqrt{b}\}$ , and thus when the manipulations leading to (2.22) are carried out with  $\chi \equiv 1$  and  $k(z, w) = \Phi(\frac{|z-w|^2}{\text{Im } z \cdot \text{Im } w})$  for some nonnegative function  $\Phi$  as in Lemma 2.1 we obtain in particular  $\int_D k(z, Tz) d\mu(z) < \infty$ . But the integral is seen to equal  $\int_{\sqrt{b}}^{\infty} \int_0^{\infty} \Phi(u) \frac{du}{4\sqrt{u}} \frac{dy}{y}$  (imitating [17, pp. 25–26]) and this is divergent for any non-negative function  $\Phi \neq 0$ . This is a contradiction.

Since  $Z_\Gamma(T)$  is discrete and contains hyperbolic elements, there is a smallest number  $\gamma_0 > 1$  with  $\begin{pmatrix} \gamma_0 & 0 \\ 0 & 1 \end{pmatrix} \in Z_\Gamma(T)$ . Using also the fact that  $T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \in Z_\Gamma(T)$ , we find that  $Z_\Gamma(T) = \{ \begin{pmatrix} \pm \gamma_0^n & 0 \\ 0 & 1 \end{pmatrix} \mid n \in \mathbb{Z} \}$  or  $Z_\Gamma(T) = \{ \begin{pmatrix} \pm \gamma_0^n & 0 \\ 0 & 1 \end{pmatrix} \mid n \in \mathbb{Z} \} \cup \{ \begin{pmatrix} 0 & \pm b \gamma_0^n \\ 1 & 0 \end{pmatrix} \mid n \in \mathbb{Z} \}$  for some

$b > 0$ . Now by similar computations as in [17, pp. 25–26] we see that the contribution from  $\{T\}$  is indeed as line 2 of (2.21).

The cases when  $T$  is hyperbolic or a glide reflection are treated by similar discussions, and are somewhat easier. Finally for the case of  $T$  elliptic see [17, pp. 351, 449–450] (however note that in our setting  $Z_\Gamma(T)$  may also contain reflections, in the special case  $T^2 = I$ ).  $\square$

*Remark 2.3.* It is convenient for applications to rewrite (2.21) using instead the group  $\overline{G} := \{g \in \mathrm{GL}(2, \mathbb{R}) \mid \det g = \pm 1\}$  and the preimages  $\overline{\Gamma}^+, \overline{\Gamma}, \overline{M} \subset \overline{G}$  of  $\Gamma^+, \Gamma$  and  $M$  under the natural projection map  $\pi : \overline{G} \rightarrow G = \mathrm{PGL}(2, \mathbb{R})$  ( $\pi$  is a double cover;  $G \cong \overline{G}/\{\pm I\}$ ). By a case-by-case comparison of the  $\Gamma$ -conjugacy classes in  $M$  and the  $\overline{\Gamma}^+$ -conjugacy classes in  $\overline{M}$  and the corresponding centralizers one shows that lines 2 and 3 in (2.21) equal

$$(2.23) \quad \frac{1}{4} \sum_{\substack{\{T\}_+ \subset \overline{M}, \\ \text{non-ell}}} \frac{\chi(T) \cdot \log N(T_0)}{N(T)^{1/2} - \det T \cdot N(T)^{-1/2}} g(\log N(T)) \\ + \frac{1}{4} \sum_{\substack{\{T\}_+ \subset \overline{M}, \\ \text{elliptic}}} \frac{\chi(T)}{|Z_{\overline{\Gamma}^+}(T)| \cdot \sin \theta(T)} \cdot \int_{-\infty}^{\infty} \frac{e^{-2\theta(T)r}}{1 + e^{-2\pi r}} \cdot h(r) dr,$$

where the sums are taken over all  $\overline{\Gamma}^+$ -conjugacy classes in  $\overline{M}$ , and for non-elliptic  $T \in \overline{M}$  we choose  $T_0 \in \overline{M}$  hyperbolic so that  $Z_{\overline{\Gamma}^+}(T) = \{\pm T_0^k \mid k \in \mathbb{Z}\}$ . (Also, we write  $\chi(T)$  in place of  $\chi(\pi(T))$ , and similarly for  $N(T)$  and  $\theta(T)$ .)

We now turn to the remaining integral in (2.19).

**Lemma 2.4.** *The set  $\Gamma - M_I$  contains only reflections and parabolic elements, and for each reflection  $T \in \Gamma - M_I$  the centralizer  $Z_{\Gamma^+}(T)$  is finite and both fixpoints of  $T$  are cusps. The integral in the second line of (2.19) equals:*

$$\lim_{Y \rightarrow \infty} \left( \sum_{\substack{\{T\}_+ \subset \Gamma - M_I \\ \text{reflection}}} \frac{\chi(T)}{2|Z_{\Gamma^+}(T)|} \int_{(*)_{T,Y}} k(z, Tz) d\mu(z) + \sum_{\substack{\{T\}_+ \subset \Gamma - M_I \\ \text{parabolic}}} \frac{\chi(T)}{2} \int_{(*)_{T,Y}} k(z, Tz) d\mu(z) \right. \\ \left. - \frac{1}{2} \int_{\mathcal{F}_Y} H(z, z, \chi) d\mu(z) \right).$$

Here the  $\{T\}_+$  sums are taken over sets of representatives of the distinct  $\Gamma^+$ -conjugacy classes in  $\Gamma - M_I$  (not  $\Gamma$ -conjugacy classes!), and if  $T$  is a reflection with fixpoints  $V_1\eta_k, V_2\eta_\ell$  ( $V_1, V_2 \in \Gamma^+, k, \ell \in \{1, \dots, \kappa\}$ ) then  $(*)_{T,Y}$  is the region defined by

$$(*)_{T,Y} := \mathcal{H} - V_1 N_k^{-1} \{z \mid \mathrm{Im} z > Y\} - V_2 N_\ell^{-1} \{z \mid \mathrm{Im} z > Y\};$$

if  $T$  is parabolic with fixpoint  $V_1\eta_k$  ( $V_1 \in \Gamma^+, k \in \{1, \dots, \kappa\}$ ) then  $(*)_{T,Y}$  is defined by

$$(*)_{T,Y} := V_1 N_k^{-1} [(0, 1) \times (0, Y)].$$

*Proof.* If  $T \in \Gamma - M_I$  were hyperbolic then we would have  $T \in \Gamma^+$ , contradicting the well-known fact that no hyperbolic element in  $\Gamma^+$  can fix a  $\Gamma^+$ -cusp. Now  $T \in \Gamma - M_I$  cannot be a glide reflection either, since then  $T^2 \in \Gamma^+$  would be hyperbolic and fix a  $\Gamma^+$ -cusp. This proves our first claim. If  $T \in \Gamma - M_I$  is a fixed reflection then by a similar argument one proves that  $Z_\Gamma(T)$  cannot contain any hyperbolic elements or glide reflections; hence as in the proof of Proposition 2.2,  $Z_\Gamma(T)$  (and thus  $Z_{\Gamma^+}(T)$ ) is finite.



Recall the definition of  $\mathcal{F}$ ,  $B_0$ ,  $\mathcal{F}_Y$  and  $\mathcal{C}_{jY}$  from p. 4; since  $[\Gamma : \Gamma^+] = 2$ , the integral in the second line of (2.19) can be expressed as  $\frac{1}{2} \int_{\mathcal{F}_Y} \cdots + \frac{1}{2} \sum_{j=1}^{\kappa} \int_{\mathcal{C}_{jY}} \cdots$ , with arbitrary  $Y \geq B_0$ . In view of Lemma 2.1 (both statements) we may then change the order of summation and integration to obtain

$$(2.24) \quad \frac{1}{2} \sum_{T \in \Gamma - M_I} \chi(T) \int_{\mathcal{F}_Y} k(z, Tz) d\mu(z) + \frac{1}{2} \sum_{j=1}^{\kappa} \chi(T) \sum_{T \in (\Gamma - M_I) - F_j} \int_{\mathcal{C}_{jY}} k(z, Tz) d\mu(z) \\ - \frac{1}{2} \int_{\mathcal{F}_Y} H(z, z, \chi) d\mu(z) + \frac{1}{2} \sum_{j=1}^{\kappa} \int_{\mathcal{C}_{jY}} \left( \sum_{T \in F_j - \{I\}} \chi(T) k(z, Tz) - H(z, z, \chi) \right) d\mu(z).$$

In fact it follows that the expressions in the first line are absolutely convergent for any choice of  $\Phi(t)$  as in Lemma 2.1.

Now let  $T \in \Gamma - M_I$  be a fixed reflection. If  $A$  runs through  $\Gamma^+$  then  $A^{-1}TA$  visits every element in the  $\Gamma^+$ -conjugacy class  $\{T\}_+ \subset \Gamma - M_I$  exactly  $|Z_{\Gamma^+}(T)|$  times. Hence the contribution from the conjugacy class  $\{T\}_+$  to the first line of (2.24) equals  $\frac{\chi(T)}{2|Z_{\Gamma^+}|} \int_D k(z, Tz) d\mu(z)$ , where  $D$  is the union of the regions  $A(\mathcal{F}_Y)$  for all  $A \in \Gamma^+$  and the regions  $A(\mathcal{C}_{jY})$  over all  $\langle A, j \rangle \in \Gamma^+ \times \{1, \dots, \kappa\}$  except those with  $A^{-1}TA \in F_j$ ; note that this is a disjoint union, disregarding sets of measure 0. We first assume that *both* fixpoints of  $T$  are cusps, say  $V_1\eta_k$  and  $V_2\eta_\ell$  ( $V_1, V_2 \in \Gamma^+$ ,  $k, \ell \in \{1, \dots, \kappa\}$ ). We then have  $A^{-1}TA \in F_j$  if and only if  $[A \in V_1[T_k], j = k]$  or  $[A \in V_2[T_\ell], j = \ell]$ , and hence we find that  $D = (*)_{T,Y}$ , as in the formulation of the lemma. If, on the other hand, only *one* fixpoint of  $T$  would be a cusp (say  $V_1\eta_k$ ) while the other fixpoint is not a cusp, then we would obtain  $D = \mathcal{H} - V_1N_k^{-1}\{z \mid \text{Im } z > Y\}$  and by an auxiliary conjugation one then sees  $\int_D k(z, Tz) d\mu(z) = \int_{\{\text{Im } z < Y\}} k(z, -\bar{z}) d\mu(z)$ . However this integral is easily seen to be divergent for every non-negative choice of  $\Phi \neq 0$ , contradicting our earlier remark about absolute convergence.

Similar manipulations also show that the contribution from the parabolic conjugacy classes  $\{T\}_+ \in \Gamma - M_I$  to the first line in (2.24) is as in the formula in the lemma. (Cf. also [37, p. 25].) Finally note that each  $\mathcal{C}_{jY}$ -integral in the second line of (2.24) tends to 0 as  $Y \rightarrow \infty$ ; this is clear from (2.20) and Lemma 2.1. This completes the proof of the lemma.  $\square$

We will now enumerate the  $\Gamma^+$ -conjugacy classes which appear in Lemma 2.4 in a way which is well suited for applications to explicit cases. For each  $j \in \{1, \dots, \kappa\}$  such that  $k(j) = j$  and each  $v \in \mathbb{Z}$  we define

$$(2.25) \quad T_{jv} := N_j^{-1} \begin{pmatrix} -1 & x_j + v \\ 0 & 1 \end{pmatrix} N_j = U_j^{-1} V T_j^v \in \Gamma.$$

(The last identity follows from (2.5).) Note that  $T_{jv}$  is a reflection fixing  $\eta_j$ ; in particular  $T_{jv} \in \Gamma - M_I$ . It is clear from the proofs of Proposition 2.2 and Lemma 2.4 that for each reflection  $T \in \Gamma - M$  we have  $|Z_{\Gamma^+}(T_{jv})| = 1$  or  $= 2$ .

**Lemma 2.5.** *Let  $H$  be the set of all pairs  $\langle j, v \rangle$  such that  $k(j) = j$  and  $v \in \{0, 1\}$ . Then for each reflection  $T \in \Gamma - M$  there is some pair  $\langle j, v \rangle \in H$  such that  $T$  is  $\Gamma^+$ -conjugate to  $T_{jv}$ . For each  $\langle j, v \rangle \in H$  there are exactly  $2|Z_{\Gamma^+}(T_{jv})|^{-1}$  pairs  $\langle j', v' \rangle \in H$  such that  $T_{jv}$  and  $T_{j'v'}$  are  $\Gamma^+$ -conjugate.*

*Proof.* Let  $T \in \Gamma - M$  be a reflection, and let  $\eta$  be one of  $T$ 's two fixpoints. By Lemma 2.4,  $\eta$  is a  $\Gamma^+$ -cusp; hence  $\eta = A_1\eta_j$  for some  $j \in \{1, \dots, \kappa\}$ ,  $A_1 \in \Gamma^+$ . Using  $A_1^{-1}TA_1 \in V^{-1}\Gamma^+$  and (2.5) we see that there is some  $B \in \Gamma^+$  such that  $A_1^{-1}TA_1 = N_j^{-1} \begin{pmatrix} -1 & x_j \\ 0 & 1 \end{pmatrix} N_{k(j)}B$ . But

$A_1^{-1}TA_1(\eta_j) = \eta_j$ ; hence  $N_{k(j)}B(\eta_j) = \begin{pmatrix} -1 & x_j \\ 0 & 1 \end{pmatrix} N_j(\eta_j) = \infty$ , and thus  $k(j) = j$  and  $B = T_j^n$  for some  $n \in \mathbb{Z}$ . Writing  $n = v - 2m$  with  $v \in \{0, 1\}$ ,  $m \in \mathbb{Z}$ , a simple computation now shows  $(A_1T_j^m)^{-1}TA_1T_j^m = T_{jv}$ . Hence  $T$  is  $\Gamma^+$ -conjugate to  $T_{jv}$ , and  $\langle j, v \rangle \in H$ .

We leave the details of the uniqueness discussion to the reader.  $\square$

The explicit evaluation of the ‘‘reflection integral’’ in Lemma 2.4 is as follows (cf. also [42, pp. 135–136]). We keep  $Y \geq B_0$ . Note that  $T_{jv}$  has one fixpoint  $\eta_j$ ; write the other fixpoint as  $V_2\eta_k$  (as before). Define the number  $\mathbf{c}_{jv} > 0$  by the relation

$$(2.26) \quad N_j V_2 N_k^{-1} = \begin{pmatrix} a & b \\ \mathbf{c}_{jv} & d \end{pmatrix}, \quad \text{with } \det \begin{pmatrix} a & b \\ \mathbf{c}_{jv} & d \end{pmatrix} = 1.$$

By an auxiliary conjugation one sees  $\int_{(*)_{T_{jv}, Y}} k(z, T_{jv}z) d\mu(z) = \int_{D_Y} k(z, -\bar{z}) d\mu(z)$  where  $D_Y$  is the whole upper half plane minus the two horoballs  $\{y > Y\}$  and  $\{|z - \frac{i}{2c^2Y}| < \frac{1}{2c^2Y}\}$  with  $c = \mathbf{c}_{jv}$ . The integral is further evaluated by changing variables to  $z = ue^{i\varphi}$  and  $t = 4 \cot^2 \varphi$  and using [16, 4.2923, 3.5416] and (2.11), (2.12) and (2.13). We now use Lemma 2.5, and note that  $\sum_{\langle j, v \rangle \in H} \chi(T_{jv}) = 2 \sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0})$ , for if  $k(j) = j$  then  $\chi(T_{j1}) = \chi(T_{j0})\chi(T_j)$  and also  $\chi(T_j) \in \{1, -1\}$  (since  $T_{j0}T_jT_{j0}^{-1} = T_j^{-1}$ ). Hence the full ‘‘reflection sum’’ in Lemma 2.4 equals:

$$(2.27) \quad \left( \sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0}) \right) \left\{ \frac{1}{2}g(0) \log Y + \frac{1}{8}h(0) + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \left( \frac{\Gamma'(\frac{1}{2} + ir)}{\Gamma(\frac{1}{2} + ir)} - \frac{\Gamma'(1 + ir)}{\Gamma(1 + ir)} \right) dr \right\} \\ + \frac{g(0)}{4} \sum_{\langle j, v \rangle \in H} \chi(T_{jv}) \log \mathbf{c}_{jv}.$$

The parabolic sum in Lemma 2.4 is exactly as in the usual trace formula; as in [18, p. 313 (4.13)] we find that the parabolic sum in Lemma 2.4 equals, as  $Y \rightarrow \infty$ :

$$(2.28) \quad \frac{|C_{\Gamma, \chi}|}{2} \left\{ g(0) \log Y - g(0) \log 2 + \frac{1}{4}h(0) - \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'(1 + ir)}{\Gamma(1 + ir)} dr \right\} \\ + \frac{g(0)}{2} \sum_{\substack{1 \leq j \leq \kappa \\ j \notin C_{\Gamma, \chi}}} \log \frac{1}{|1 - \chi(T_j)|} + O(Y^{-\frac{1}{2}}).$$

It remains to evaluate the  $H(z, z, \chi)$ -integral in Lemma 2.4. Let  $\Phi(s) = (\varphi_{jk}(s))_{j, k \in C_{\Gamma, \chi}}$  be the scattering matrix for  $\Gamma^+, \chi$  ([18, p. 281 (Def. 3.8)]). We point out the following relations:

$$(2.29) \quad \varphi_{j\ell}(s) = \chi(U_j^{-1}U_\ell)\varphi_{k(j)k(\ell)}(s), \quad \forall j, \ell \in C_{\Gamma, \chi};$$

$$(2.30) \quad \chi(V^{-1}U_j) = \chi(VU_{k(j)}^{-1}), \quad \forall j \in C_{\Gamma, \chi}.$$

Relation (2.29) follows, for fixed generic  $s \in \mathbb{C}$ , from the asymptotic formula for  $E_j(z, s, \chi)$  in the cusp  $\eta_\ell$  ([18, pp. 280, 297(F)]) together with (2.8) and (2.5). Relation (2.30) follows directly from (2.6), since  $\chi(T_j) = 1$  for  $j \in C_{\Gamma, \chi}$ .

For any  $j, \ell \in C_{\Gamma, \chi}$  we have the following Maass-Selberg identity, whose proof is a straightforward generalization of [18, pp. 153–155, 200–201(13.15)] (see [37, §7] for a detailed proof

in the case  $\chi \equiv 1$ ):

(2.31)

$$\int_{\mathcal{F}_Y} E_j(z, \frac{1}{2} + it, \chi) \overline{E_\ell(z, \frac{1}{2} + it, \chi)} d\mu(z) = 2\delta_{j\ell} \log Y - \sum_{m \in C_{\Gamma, \chi}} \varphi'_{jm}(\frac{1}{2} + it) \varphi_{m\ell}(\frac{1}{2} - it) + \frac{\varphi_{j\ell}(\frac{1}{2} - it)Y^{2it} - \varphi_{j\ell}(\frac{1}{2} + it)Y^{-2it}}{2it} + O(\omega(t)e^{6|t|-4\pi Y}),$$

uniformly over all  $Y \geq B_0$  and all  $t \in \mathbb{R} - \{0\}$ . Here  $\omega(t)$  is an even function which only depends on  $\Gamma$  and which satisfies  $\omega(t) \geq 1$  and  $\int_{-T}^T \omega(t) dt = O(T^2)$  as  $T \rightarrow \infty$  ([18, pp. 161, 299(line 14), 315(ii)]).

We will need an analog of (2.31) for the  $\Gamma$ -Eisenstein series  $E^\Gamma(z, s, \chi)$  (cf. (2.9), (2.10)). We first introduce some more notation. We fix, once and for all, a subset  $R_{\Gamma, \chi} \subset C_{\Gamma, \chi}$  such that

$$\forall j \in C_{\Gamma, \chi} : \quad \begin{cases} k(j) \neq j \implies R_{\Gamma, \chi} \text{ contains exactly one of } j, k(j); \\ k(j) = j \implies [j \in R_{\Gamma, \chi} \text{ iff. } \chi(V^{-1}U_j) = 1]. \end{cases}$$

(Recall that if  $k(j) = j$  then  $\chi(V^{-1}U_j) = \chi(T_{j0}^{-1}) = \pm 1$ ) Also, for all  $j, \ell \in R_{\Gamma, \chi}$  we define

$$(2.32) \quad \varepsilon_\ell = \begin{cases} 2 & \text{if } k(\ell) = \ell \\ 1 & \text{if } k(\ell) \neq \ell, \end{cases} \quad \text{and} \quad \varphi_{j\ell}^\Gamma(s) = \varepsilon_\ell^{-1} \left( \varphi_{j\ell}(s) + \chi(V^{-1}U_\ell) \varphi_{j, k(\ell)}(s) \right).$$

As a partial (and sketchy!) motivation for these definitions, note that if  $f \in C^\infty(\mathcal{H})$  is any  $\Gamma, \chi$ -invariant eigenfunction with  $\Delta$ -eigenvalue  $s(1-s)$ ,  $\text{Re } s > 1$  and of modest growth in each cusp (of  $\Gamma^+ \backslash \mathcal{H}$ ), then  $f(z)$  decays exponentially in all cusps  $\eta_j$  with  $j \notin C_{\Gamma, \chi}$  or  $[j \in C_{\Gamma, \chi}, k(j) = j, \chi(V^{-1}U_j) = -1]$  (see [18, Ch. 6 (Prop. 4.10)]; in the second case also note that by (2.5),  $f$  is odd with respect to a geodesical axis through the cusp  $\eta_j$ ). Using [18, pp. 69(bottom)–71, 282 (item 2)] and the symmetry between the cusps  $\eta_j$  and  $\eta_{k(j)}$  (see (2.5)) one then proves  $f(z) \equiv \sum_{\ell \in R_{\Gamma, \chi}} \varepsilon_\ell^{-1} a_\ell E_\ell^\Gamma(z, s, \chi)$ , where the constants  $a_\ell \in \mathbb{C}$  ( $\ell \in R_{\Gamma, \chi}$ ) are determined by  $f(N_\ell^{-1}(x + iy)) \sim a_\ell y^s$  as  $y \rightarrow \infty$ . In particular, in this way one proves for each  $j \in R_{\Gamma, \chi}$  (first for  $\text{Re } s < 0$  and then for generic  $s \in \mathbb{C}$  by meromorphic continuation):  $E_j^\Gamma(z, 1-s, \chi) \equiv \sum_{\ell \in R_{\Gamma, \chi}} \varphi_{j\ell}^\Gamma(1-s) E_\ell^\Gamma(z, s, \chi)$ , where  $\varphi_{j\ell}^\Gamma(s)$  is as in (2.32). Applying this formula twice and using the uniqueness of the constants  $a_\ell$  in the above expansion  $f(z) \equiv \sum_{\ell \in R_{\Gamma, \chi}} \varepsilon_\ell^{-1} a_\ell E_\ell^\Gamma(z, s, \chi)$ , we obtain (if  $R_{\Gamma, \chi} \neq \emptyset$ )

$$(2.33) \quad \Phi^\Gamma(s) \Phi^\Gamma(1-s) = I, \quad \text{where } \Phi^\Gamma(s) := [\varphi_{j\ell}^\Gamma(s)]_{j, \ell \in R_{\Gamma, \chi}}.$$

(Thus the matrix  $\Phi^\Gamma(s)$  is a good  $\Gamma$ -analog of the usual  $\Phi(s)$ , cf. [18, p. 281].) In retrospect the last two relations may alternatively be verified directly from our definitions together with [18, p. 283 (3.13)], by direct but rather tedious computations.

With this notation in place we have, for any  $j, \ell \in R_{\Gamma, \chi}$ , the following version of the Maass–Selberg identity:

$$(2.34) \quad \int_{\mathcal{F}_Y} E_j^\Gamma(z, \frac{1}{2} + it, \chi) \overline{E_\ell^\Gamma(z, \frac{1}{2} + it, \chi)} d\mu(z) = 2\varepsilon_\ell \left( 2\delta_{j\ell} \log Y - \sum_{m \in R_{\Gamma, \chi}} (\varphi_{jm}^\Gamma)'(\frac{1}{2} + it) \varphi_{m\ell}^\Gamma(\frac{1}{2} - it) + \frac{\varphi_{j\ell}^\Gamma(\frac{1}{2} - it)Y^{2it} - \varphi_{j\ell}^\Gamma(\frac{1}{2} + it)Y^{-2it}}{2it} \right) + O(\omega(t)e^{6|t|-4\pi Y}),$$

uniformly over all  $Y \geq B_0$  and all  $t \in \mathbb{R} - \{0\}$ . The proof of (2.34) is (again) a tedious but direct computation straight from the definitions (2.10) and (2.32), using (2.31) and the relations (2.29), (2.30). (A more illuminating proof of (2.34) – but requiring more work – is to carry out the Maass–Selberg computation from scratch working on the space  $\Gamma \backslash \mathcal{H}$ .)

It follows from (2.10) that  $E_{k(j)}^\Gamma(z, s, \chi) = \chi(V^{-1}U_j)E_j^\Gamma(z, s, \chi)$ . Hence the formula for  $H(z, w, \chi)$ , (2.16), can be rewritten as

$$H(z, w, \chi) = \frac{1}{4\pi} \sum_{j \in R_{\Gamma, \chi}} \varepsilon_j^{-1} \int_{-\infty}^{\infty} h(t) E_j^\Gamma(z, \frac{1}{2} + it, \chi) \overline{E_j^\Gamma(w, \frac{1}{2} + it, \chi)} dt.$$

(In particular  $H(z, w, \chi) \equiv 0$  if  $R_{\Gamma, \chi} = \emptyset$ .) Now to evaluate  $\frac{1}{2} \int_{\mathcal{F}_Y} H(z, z, \chi) d\mu(z)$  we may change order of integration (this is justified by absolute convergence of the double integral; recall  $h(t) = O(e^{-15|t|})$  and cf. [18, p. 301(a,d)]), and then apply (2.34). The contribution from the second line in (2.34) is treated as in [18, p. 202]. To simplify the answer we define  $\varphi^\Gamma(s) := \det \Phi^\Gamma(s)$  and note that (using (2.33))

$$(2.35) \quad \frac{(\varphi^\Gamma)'(s)}{\varphi^\Gamma(s)} = \text{Tr}[(\Phi^\Gamma)'(s) \cdot \Phi^\Gamma(s)^{-1}] = \sum_{j, m \in R_{\Gamma, \chi}} (\varphi_{jm}^\Gamma)'(s) \cdot \varphi_{mj}^\Gamma(1-s).$$

We thus obtain:

$$(2.36) \quad \begin{aligned} \frac{1}{2} \int_{\mathcal{F}_Y} H(z, z, \chi) d\mu(z) &= |R_{\Gamma, \chi}| \cdot g(0) \log Y - \frac{1}{4\pi} \int_{-\infty}^{\infty} h(t) \cdot \frac{(\varphi^\Gamma)'(\frac{1}{2} + it)}{\varphi^\Gamma(\frac{1}{2} + it)} dt \\ &\quad + \frac{1}{4} h(0) \cdot \text{Tr} \Phi^\Gamma(\frac{1}{2}) + O(Y^{1-2\beta}), \quad \text{as } Y \rightarrow \infty, \end{aligned}$$

where  $\beta > \frac{1}{2}$  is a constant which only depends on  $\Gamma$ . Finally, using the definition and our remarks about  $R_{\Gamma, \chi}$ , one easily checks that

$$|R_{\Gamma, \chi}| = \frac{1}{2} \sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0}) + \frac{1}{2} |C_{\Gamma, \chi}|.$$

Hence the  $\log Y$  terms cancel each other when adding (2.27) and (2.28) and subtracting (2.36) (as in Lemma 2.4), and hence we may compute the limit as  $Y \rightarrow \infty$ . Combining this with (2.19) and Proposition 2.2 we obtain, at last, *the explicit trace formula for  $\Gamma, \chi$* :

$$(2.37) \quad \begin{aligned} \sum_{n \geq 1} h(r_n) &= \frac{\mu(\Gamma \backslash \mathcal{H})}{4\pi} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr \\ &\quad + \sum_{\substack{\{T\} \subset M, \\ \text{non-ell}}} \frac{\log N(T_0)}{[Z_\Gamma(T) : [T_0]]} \cdot \frac{\chi(T) \cdot g(\log N(T))}{N(T)^{1/2} - \text{sgn}(\det T) \cdot N(T)^{-1/2}} \\ &\quad + \sum_{\substack{\{T\} \subset M, \\ \text{elliptic}}} \frac{\chi(T)}{2|Z_\Gamma(T)| \cdot \sin \theta(T)} \cdot \int_{-\infty}^{\infty} \frac{e^{-2\theta(T)r}}{1 + e^{-2\pi r}} \cdot h(r) dr \\ &\quad + \left( \sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0}) \right) \left\{ \frac{1}{8} h(0) + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \left( \frac{\Gamma'(\frac{1}{2} + ir)}{\Gamma(\frac{1}{2} + ir)} - \frac{\Gamma'(1 + ir)}{\Gamma(1 + ir)} \right) dr \right\} \end{aligned}$$

$$\begin{aligned}
 & + \frac{g(0)}{4} \sum_{(j,v) \in H} \chi(T_{jv}) \log \mathbf{c}_{jv} + \frac{g(0)}{2} \sum_{\substack{1 \leq j \leq \kappa \\ j \notin C_{\Gamma, \chi}}} \log \frac{1}{|1 - \chi(T_j)|} \\
 & + \frac{|C_{\Gamma, \chi}|}{2} \left\{ \frac{1}{4} h(0) - g(0) \log 2 - \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'(1+ir)}{\Gamma(1+ir)} dr \right\} \\
 & + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \cdot \frac{(\varphi^\Gamma)'(\frac{1}{2} + ir)}{\varphi^\Gamma(\frac{1}{2} + ir)} dr - \frac{1}{4} h(0) \cdot \text{Tr } \Phi^\Gamma(\frac{1}{2}).
 \end{aligned}$$

Here the set  $M$  was defined in (2.18), and  $Z_\Gamma(T), N(T), \theta(T), T_0$  were defined in and before Proposition 2.2. The set  $C_{\Gamma, \chi}$  is  $C_{\Gamma, \chi} = \{j \in \{1, \dots, \kappa\} \mid \chi(T_j) = 1\}$  and the set  $H$  is  $H = \{(j, v) \in \{1, \dots, \kappa\} \times \{0, 1\} \mid k(j) = j\}$  as in Lemma 2.5. We have  $T_{jv} = N_j^{-1} \begin{pmatrix} -1 & x_j + v \\ 0 & 1 \end{pmatrix} N_j$  (see (2.25)). The numbers  $\mathbf{c}_{jv} > 0$  were defined in (2.26), and the matrix  $\Phi^\Gamma(s)$  was defined in (2.33); also  $\varphi^\Gamma(s) := \det \Phi^\Gamma(s)$ ; if  $R_{\Gamma, \chi} = \emptyset$  then the last line in (2.37) is to be omitted.

We recall that an alternative form for the second and third lines in (2.37) was given in Remark 2.3. We also note that if  $\Gamma$  is cocompact ( $\Leftrightarrow \Gamma^+$  is cocompact) then the last four lines in (2.37) are to be omitted (for in this case we have  $M = \Gamma$  and hence the trace formula is given already by Proposition 2.2).

Finally, by a standard approximation argument we may extend the class of admissible test functions  $h(r)$ :

**Theorem 2.** *Let  $h(r)$  be any function satisfying the following hypotheses:*

- (a)  $h(r)$  is analytic on  $|\text{Im } r| \leq \frac{1}{2} + \delta$  for some  $\delta > 0$ ;
- (b)  $h(-r) = h(r)$ ;
- (c)  $|h(r)| \leq A(1 + |\text{Re } r|)^{-2-\delta}$ .

Set

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) e^{-iru} dr, \quad u \in \mathbb{R}.$$

Then the formula (2.37) holds, where the various sums and integrals are all absolutely convergent.

*Proof.* This is very similar to [18, pp. 209-210 (Thm. 13.8)]. To obtain good majorants for the  $((\varphi^\Gamma)'/\varphi^\Gamma)$ -integral one uses (2.35) and computations similar to the proof of (2.34) to show

$$\begin{aligned}
 \frac{(\varphi^\Gamma)'(\frac{1}{2} + it)}{\varphi^\Gamma(\frac{1}{2} + it)} &= \frac{1}{2} \sum_{j, m \in C_{\Gamma, \chi}} \varphi'_{jm}(\frac{1}{2} + it) \overline{\varphi_{jm}(\frac{1}{2} + it)} \\
 &+ \frac{1}{2} \sum_{j \in C_{\Gamma, \chi}} \chi(V^{-1}U_j) \sum_{m \in C_{\Gamma, \chi}} \varphi'_{jm}(\frac{1}{2} + it) \overline{\varphi_{k(j), m}(\frac{1}{2} + it)}.
 \end{aligned}$$

This can now be bounded using [37, Lemma 8.2] (which is easily generalized from  $\chi \equiv 1$  to general  $\chi$ ).  $\square$

*Remark 2.6.* The situation which we have considered in the present section contains as a special case the situation in Venkov [42, §6.5], which in our notation corresponds to taking  $\chi = \text{sgn} \circ \det : \Gamma \rightarrow \{1, -1\}$  and special choices of  $\Gamma$ ; in particular, note that for any  $\Gamma = \Gamma_M^0$  as in [42, p. 118(top)] we have  $k(j) = j$  for all  $j \in \{1, \dots, \kappa\}$ . (This is not true

for general cofinite subgroups  $\Gamma \subset G$ ; cf., e.g., [40].) For these choices of  $\Gamma, \chi$  one has  $C_{\Gamma, \chi} = \{1, \dots, \kappa\}$  and  $R_{\Gamma, \chi} = \emptyset$  using our notation, and hence the last line in (2.37) is not present. One easily checks that (2.37) indeed specializes to the formula [42, (6.5.40)], after the latter has been slightly modified in its next to last line to correct for two minor errors (note that these mistakes were repeated in [5]): Firstly, there may exist reflections  $T \in \Gamma$  which have trivial  $\Gamma^+$ -centralizer  $Z_{\Gamma^+}(T) = \{I\}$ , contrary to what is assumed in [42, (6.5.23)]. (Take e.g.  $M$  in [42, p. 117(bottom)] to be  $M = \{z \in \mathcal{H} \mid |z| \geq 1, |\operatorname{Re} z| < \frac{1}{2}\}$ ; this leads to  $\Gamma = \Gamma_M^0 = \{T \in \operatorname{PGL}(2, \mathbb{Z}) \mid \operatorname{sgn}(\det(T)) = h(T \bmod 2)\}$  where  $h$  is the unique epimorphism  $h : \operatorname{PGL}(2, \mathbb{Z}/2\mathbb{Z}) \cong S_3 \rightarrow \{1, -1\}$ . One now finds  $Z_{\Gamma^+}(T) = \{I\}$  for the reflection  $T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \in \Gamma$ .) Secondly, contrary to what is claimed in [42, below (6.5.12)], the  $\Gamma^+$ -centralizer  $Z_{\Gamma^+}(T)$  of a reflection  $T$  may contain both hyperbolic and elliptic elements. (For a concrete example, let  $\Gamma = \Gamma_M^0$  be the group generated by the reflections in the sides of any compact hyperbolic triangle  $M$  with angles  $\pi/n_j$ , where  $n_1, n_2, n_3 \geq 2$  and  $n_1^{-1} + n_2^{-1} + n_3^{-1} < 1$ . Then if  $n_1$  is even,  $\Gamma^+$  contains the elliptic element  $E$  which rotates angle  $\pi$  about the corresponding vertex; if  $T \in \Gamma$  is the reflection in one of the adjacent sides then  $E \in Z_{\Gamma^+}(T)$ , but  $Z_{\Gamma^+}(T)$  also contains an infinite hyperbolic cyclic subgroup, since  $\Gamma$  is cocompact.)

**2.2. Trace formula for  $\Gamma_0(N), \chi$ , with  $N$  squarefree.** Throughout this subsection we will use the convention that all matrix representatives for elements in  $G = \operatorname{PGL}(2, \mathbb{R})$  are taken to have determinant 1 or  $-1$ . Let  $N \in \mathbb{Z}^+$  be squarefree and set  $\Gamma^+ = \Gamma_0(N) \subset G^+$ . Let  $V = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \in G$  and note that  $V^2 = I$ ,  $V\Gamma^+V^{-1} = \Gamma^+$ ; hence  $\Gamma = \Gamma_0^\pm(N) := \langle \Gamma^+, V \rangle \subset G$  is a supergroup of  $\Gamma^+$  of index 2. Explicitly we have

$$(2.38) \quad \Gamma = \Gamma_0^\pm(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{Z}, ad - bc = \pm 1, N|c \right\}.$$

Let  $\chi$  be an even Dirichlet character modulo  $N$ . Let  $q(\chi)$  be its conductor. For any divisor  $A \mid N$  we will let  $\chi_A$  denote the character modulo  $A$  defined by  $\chi_A(x) := \chi(y)$  whenever  $y \equiv x \pmod{A}$  and  $y \equiv 1 \pmod{N/A}$ . It follows that  $\chi = \chi_A \cdot \chi_{N/A}$  for all divisors  $A \mid N$  (note that  $A$  and  $N/A$  are always relatively prime since  $N$  is squarefree). We will use the notation  $\chi^{(A)} := \chi_A \overline{\chi_{q(\chi)/A}}$  (thus  $\chi^{(A)}$  is a primitive character modulo  $q(\chi)$ ). Also, let us agree to call  $\chi$  *pure* if  $\chi_p(-1) = 1$  for every prime  $p \mid N$ .

We view  $\chi$  as a character on  $\Gamma^+ = \Gamma_0(N)$  via  $\chi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) := \chi(d)$ . This character can be extended in exactly two ways to a character of  $\Gamma = \Gamma_0^\pm(N)$ ; either by  $\chi(V) = 1$  or  $\chi(V) = -1$ . These extensions are explicitly given by  $\chi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = (ad - bc)^\varepsilon \chi(d)$ ,  $\forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0^\pm(N)$ , where  $\varepsilon \in \{0, 1\}$ . From now on we fix  $\varepsilon \in \{0, 1\}$  and let  $\chi$  denote the corresponding character of  $\Gamma_0^\pm(N)$ .

Our goal in this section is to prove the following theorem.

**Theorem 3.** *The trace formula for  $\Gamma = \Gamma_0^\pm(N)$  with  $N$  squarefree and with character  $\chi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = (ad - bc)^\varepsilon \chi(d)$  can be expressed in the following explicit form:*

$$(2.39) \quad \sum_{n \geq 1} h(r_n) = \frac{\prod_{p|N}(p+1)}{24} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr \\ + \frac{1}{2} \sum_{n \in \{-1, 1\}} n^\varepsilon \sum_{\substack{t \in \mathbb{Z} \\ \sqrt{t^2 - 4n} \notin \mathbb{Q}}} \langle \chi(\delta) \rangle_{\delta^2 - t\delta + n \equiv 0}$$

$$\begin{aligned}
& \cdot \left( \sum_{f|\ell} \mathbf{h}^+(\mathfrak{r}[f]) \cdot [\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1] \cdot \prod_{p|N} \left\{ \begin{array}{ll} 2 & \text{if } p \mid f \\ 1 + \left(\frac{d}{p}\right) & \text{if } p \nmid f \end{array} \right\} \right) \cdot A(t, n) \\
& - \frac{d(N)}{4} g(0) \left( \log(8N) + C_{\chi, \varepsilon} \cdot \log \left( \frac{Nq(\chi)}{2\pi^2} \right) + \frac{1}{2}(C_{\chi, \varepsilon} - 1) \log(N, 2) \right) \\
& - \frac{d(N)}{4\pi} \int_{-\infty}^{\infty} h(r) \left( \frac{\Gamma'(\frac{1}{2} + ir)}{\Gamma(\frac{1}{2} + ir)} + C_{\chi, \varepsilon} \frac{\Gamma'(1 + ir)}{\Gamma(1 + ir)} \right) dr + \left[ \text{If } \chi \neq 1: \frac{C_{\chi, \varepsilon} d(N)}{8} h(0) \right] \\
& + d(N) \sum_{n=1}^{\infty} \frac{\Lambda(n) \cdot \{\chi\}_{\varepsilon}(n)}{n} g(2 \log n) - \frac{d(N)}{2} \sum_{p|\frac{N}{q(\chi)}} \sum_{m=1}^{\infty} \frac{(\log p) \cdot \{\chi\}_{\varepsilon}(p^m)}{p^m} g(2 \log p^m).
\end{aligned}$$

The notation is as follows. In lines 2 and 3 of (2.39),  $d, \ell$  are implicit variables defined through  $t^2 - 4n = d\ell^2$ ,  $\ell \in \mathbb{Z}^+$ ,  $d$  a fundamental discriminant;  $\mathfrak{r}[f]$  is the order  $\mathfrak{r}[f] = \mathbb{Z} + f\omega\mathbb{Z}$  with  $\omega = \frac{1}{2}(d + \sqrt{d})$ ;  $\mathfrak{r}[f]^1$  is its norm one unit group;  $\mathbf{h}^+(\mathfrak{r}[f])$  is the narrow class number for  $\mathfrak{r}[f]$ ; and  $\langle \chi(\delta) \rangle_{\delta^2 - t\delta + n \equiv 0 \pmod{q(\chi)}}$  denotes the average of  $\chi(\delta)$  taken over all  $\delta \pmod{q(\chi)}$  satisfying  $\delta^2 - t\delta + n \equiv 0 \pmod{q(\chi)}$  (there always exist solutions to this congruence equation if  $(d/p) \neq -1$  holds for each prime  $p \mid \frac{q(\chi)}{(\ell, q(\chi))}$ ; for those  $n, t$  which do not satisfy this condition, the sum over  $f$  in line 3 vanishes anyway). Furthermore,

$$(2.40) \quad A(t, n) := \begin{cases} \frac{\log \epsilon_1}{\sqrt{t^2 - 4n}} \cdot g \left( \log \frac{(|t| + \sqrt{t^2 - 4n})^2}{4} \right) & \text{if } t^2 - 4n > 0 \\ \frac{2}{|\mathfrak{r}[1]^1| \cdot \sqrt{4 - t^2}} \cdot \int_{-\infty}^{\infty} \frac{e^{-2r \cdot \arccos(|t|/2)}}{1 + e^{-2\pi r}} h(r) dr & \text{if } t^2 - 4n < 0 \\ & (\text{thus } n = 1), \end{cases}$$

where, if  $t^2 - 4n > 0$ ,  $\epsilon_1$  is the proper fundamental unit in  $\mathbb{Q}(\sqrt{d})$ . Also, in lines 4 to 6 of (2.39),  $d(N) = \#\{\delta \in \mathbb{Z}^+ : \delta|N\}$  is the usual divisor function, and

$$(2.41) \quad C_{\chi, \varepsilon} := \begin{cases} 1 & \text{if } \chi \text{ not pure} \\ 2 & \text{if } \chi \text{ pure, } \varepsilon = 0; \\ 0 & \text{if } \chi \text{ pure, } \varepsilon = 1 \end{cases} \quad \{\chi\}_{\varepsilon}(\delta) := \frac{2}{d(q(\chi))} \sum_{\substack{A|q(\chi) \\ \chi_A(-1) = (-1)^{\varepsilon}}} \chi^{(A)}(\delta).$$

More explanation of the notation is given in the proof below.

We remark that one can easily check that the parabolic contribution in (2.39) (viz., the last three lines) agrees with [18, p. 538] when added over  $\varepsilon = 0, 1$ ; also, when added over all even Dirichlet characters  $\chi$  modulo  $N$ , so that the trace formula for  $L^2(\Gamma_1(N) \backslash \mathcal{H})$  is obtained, we have agreement with the formula in [18, p. 568 ( $N \geq 3$  and odd)].

The remainder of this section is devoted to the proof of Theorem 3.

It is well known that  $\mu(\Gamma^+ \backslash \mathcal{H}) = \frac{\pi}{3} \cdot N \prod_{p|N} (1 + \frac{1}{p})$ , cf. [29, Thm. 4.2.5(2)]; hence the first line of (2.37) translates into the first line of (2.39).

The number of cusps of  $\Gamma^+ \backslash \mathcal{H}$  is  $\kappa = 2^v = d(N)$ , where  $v$  is the number of primes dividing  $N$ , and  $d$  is the usual divisor function. An explicit list of these cusps can be given as follows [18, Ch. 11, §4]. Let  $A \in \mathbb{Z}^+$  be a divisor of  $N$ . Using the Chinese remainder theorem and

[29, Thm. 4.2.1(1)] we can then select a matrix  $W_A \in \mathrm{SL}(2, \mathbb{Z})$  such that

$$(2.42) \quad W_A \equiv \begin{cases} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & \text{mod } A \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{mod } N/A, \end{cases}$$

We take  $W_A$  to be fixed once and for all. Then define

$$(2.43) \quad \eta_A = W_A(\infty), \quad N_A = \begin{pmatrix} A^{-1/2} & 0 \\ 0 & A^{1/2} \end{pmatrix} W_A^{-1}.$$

**Lemma 2.7.**  $\{\eta_A : A|N\}$  contains exactly one representative from each cusp class of  $\Gamma_0(N)$ . Also, the fixator subgroup  $\Gamma_0(N)_{\eta_A}$  equals  $\{N_A^{-1} \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} N_A \mid m \in \mathbb{Z}\}$ .

*Proof.* See [18, p. 534, lemma 4.2]. □

Now we can use the divisors  $A$  of  $N$  as a substitute for the cusp index  $j \in \{1, \dots, \kappa\}$  from the general case, thus writing  $\eta_A, N_A, T_A$  in place of  $\eta_j, N_j, T_j$ , etc. One shows by a quick computation using (2.38), (2.42) and (2.43) that  $N_A^{-1} Y N_A \in \Gamma_0^\pm(N)$  for all  $Y \in \Gamma_0^\pm(N)$ . In particular  $N_A^{-1} V N_A \in \Gamma_0^\pm(N)$ , and since  $\det N_A^{-1} V N_A = -1$  we may thus write  $N_A^{-1} V N_A = U_A^{-1} V$  for some  $U_A \in \Gamma_0(N)$ . Using  $T_A = N_A^{-1} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} N_A$  and  $V \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  it then follows that  $VT_A V^{-1} = U_A T_A^{-1} U_A^{-1}$ . This means that in the present case the decomposition in (2.5) takes a particularly simple form:  $k(A) = A$  for all  $A$ , and we may take

$$(2.44) \quad U_A = V N_A^{-1} V N_A \in \Gamma_0(N).$$

Hence the reflection representative  $T_{Av} = N_A^{-1} \begin{pmatrix} -1 & v \\ 0 & 1 \end{pmatrix} N_A \in \Gamma_0^\pm(N)$  is defined for each  $A \mid N, v \in \mathbb{Z}$ . Writing (temporarily)  $W_A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  and using (2.43) we obtain

$$T_{Av} = W_A \begin{pmatrix} -1 & vA \\ 0 & 1 \end{pmatrix} W_A^{-1} = \begin{pmatrix} * & * \\ * & ad + bc + acvA \end{pmatrix},$$

and using (2.42) we see  $ad + bc + acvA \equiv -1 \pmod{A}$  and  $\equiv 1 \pmod{N/A}$ ; hence since  $\det T_{Av} = -1$  we have

$$(2.45) \quad \chi(T_{Av}) = (-1)^\varepsilon \cdot \chi_A(-1).$$

One also checks that every cusp is open (for if  $W_A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$  then  $T_A = W_A \begin{pmatrix} 1 & -A \\ 0 & 1 \end{pmatrix} W_A^{-1} = \begin{pmatrix} * & * \\ * & 1 - A\alpha\gamma \end{pmatrix}$ , and  $A\alpha\gamma \equiv 0 \pmod{N}$ ), that is  $C_{\Gamma, \chi} = \{A : A \mid N\}$ . Hence the sum  $\sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0})$  appearing in (2.37) can be evaluated as:

$$\sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0}) = (-1)^\varepsilon \cdot \sum_{A|N} \chi_A(-1) = (-1)^\varepsilon \cdot \prod_{p|N} (1 + \chi_p(-1)).$$

Note that  $\chi_p(-1) = 1$  or  $-1$  for each prime  $p \mid N$ . We see that

$$(2.46) \quad \sum_{\substack{j \in C_{\Gamma, \chi} \\ (k(j)=j)}} \chi(T_{j0}) = (-1)^\varepsilon \begin{cases} d(N) & \text{if } \chi \text{ pure} \\ 0 & \text{otherwise.} \end{cases}$$

Let us now determine the constants  $\mathbf{c}_{Av}$  from (2.26). Clearly, the two fixpoints of  $T_{Av} = N_A^{-1} \begin{pmatrix} -1 & v \\ 0 & 1 \end{pmatrix} N_A$  are  $N_A^{-1}(\infty) = \eta_A$  and  $N_A^{-1}(v/2)$ .



**Lemma 2.8.** *The point  $N_A^{-1}(v/2)$  is  $\Gamma_0(N)$ -equivalent to  $\eta_{A'}$ , where  $A' = N/A$  if  $2 \mid v$ , and*

$$A' = \begin{cases} N/A & \text{if } 2 \nmid N \\ 2N/A & \text{if } 2 \mid A \\ N/(2A) & \text{if } 2 \nmid A, 2 \mid N \end{cases} \quad \text{if } 2 \nmid v.$$

*Proof.* Assume  $W_A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  as before; then  $N_A^{-1}(v/2) = W_A(Av/2) = \frac{aAv+2b}{cAv+2d}$ . Hence  $A' = N/(N, (cAv+2d)/s)$ , where  $s = (aAv+2b, aAv+2d)$  (cf. [12, Lemma 2.3]). But  $s \in \{1, 2\}$ , since  $a(aAv+2b) - c(aAv+2d) = 2$ . Now the lemma follows by a case-by-case analysis.  $\square$

Now in the definition of  $\mathbf{c}_{Av}$ , (2.26), we have  $V_2 \in \Gamma^+$  and  $V_2\eta_{A'} = N_A^{-1}(v/2)$ . Hence, writing  $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = W_A^{-1}V_2W_{A'} \in \text{SL}(2, \mathbb{Z})$  and using (2.43) we obtain

$$\frac{v}{2} = \begin{pmatrix} A^{-1/2} & 0 \\ 0 & A^{1/2} \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} A^{1/2} & 0 \\ 0 & A'^{-1/2} \end{pmatrix} (\infty).$$

Thus  $\frac{\alpha}{\gamma} = Av/2$ , and since  $\alpha$  and  $\gamma$  are integers and relatively prime this gives  $\gamma = \pm 1$  if  $2 \mid Av$ , otherwise  $\gamma = \pm 2$ . Hence using (2.26), (2.43) and Lemma 2.8 we get

$$\mathbf{c}_{Av} = \sqrt{AA'} \cdot |\gamma| = \begin{cases} \sqrt{N} & \text{if } 2 \mid v \\ \sqrt{2N} & \text{if } 2 \nmid v, 2 \mid N \\ 2\sqrt{N} & \text{if } 2 \nmid v, 2 \nmid N. \end{cases}$$

From this we obtain, for each  $A \mid N$ ,

$$\sum_{v \in \{0,1\}} \log \mathbf{c}_{Av} = \begin{cases} \log(\sqrt{2}N) & \text{if } 2 \mid N \\ \log(2N) & \text{if } 2 \nmid N. \end{cases}$$

Using this and (2.46) we see that lines 4 and 5 in the general trace formula (2.37) equal in our case, if  $\chi$  is pure:

$$(2.47) \quad (-1)^\varepsilon d(N) \cdot \left\{ \frac{1}{8} h(0) + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \left( \frac{\Gamma'(\frac{1}{2} + ir)}{\Gamma(\frac{1}{2} + ir)} - \frac{\Gamma'(1 + ir)}{\Gamma(1 + ir)} \right) dr \right\} \\ + (-1)^\varepsilon \frac{d(N)}{4} \left\{ \begin{array}{ll} \log(\sqrt{2}N) & \text{if } 2 \mid N \\ \log(2N) & \text{if } 2 \nmid N \end{array} \right\} \cdot g(0).$$

If  $\chi$  is not pure then lines 4 and 5 instead give contribution 0.

We consider next the Eisenstein contribution. The following lemma gives a formula for the scattering matrix, which generalizes the formula in [18, Ch. 11, §4] for trivial character to our case of arbitrary  $\chi$ .

We assume that the prime factorization of  $N$  is  $N = \prod_{j=1}^v p_j$  where we take the primes to be ordered so that  $\chi_{p_j} \neq 1$  if and only if  $1 \leq j \leq w$  (for some constant  $w \leq v$ ); thus the conductor of  $\chi$  is  $q(\chi) = \prod_{j=1}^w p_j$ . (Note that we allow  $N = 1$ , i.e.  $v = 0$ ; note also that  $w = 0$  holds iff.  $\chi$  is trivial.) Consider the scattering matrix  $\Phi(s) = (\varphi_{A_1, A_2}(s))$  where  $A_1, A_2$  both run through the positive divisors of  $N$ . We split  $\Phi(s)$  as a block matrix  $\Phi(s) = (B_{A'_1, A'_2}(s))$ , where  $A'_1, A'_2$  run through the divisors of  $q(\chi)$ , and the blocks are the square matrices  $B_{A'_1, A'_2}(s) = (\varphi_{A_1, A_2}(s))$  where  $A_1, A_2$  run through those pairs of divisors of  $N$  which satisfy  $(A_j, q(\chi)) = A'_j$ . It now turns out that in each row  $A'_1$  there is exactly one non-zero block  $B_{A'_1, A'_2}(s)$ , and this block can be determined explicitly:

**Lemma 2.9.** *We have  $B_{A'_1, A'_2}(s) = 0$  whenever  $A'_1 A'_2 \neq q(\chi)$ . In the case  $A'_1 A'_2 = q(\chi)$  we have*

$$B_{A'_1, A'_2}(s) = \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{L(2s - 1, \chi^{(A'_1)})}{L(2s, \chi^{(A'_1)})} \cdot q(\chi)^{-s} \cdot \mathcal{N}_{p_{w+1}}^{(A'_1)}(s) \otimes \mathcal{N}_{p_{w+2}}^{(A'_1)}(s) \otimes \cdots \otimes \mathcal{N}_{p_v}^{(A'_1)}(s),$$

where

$$\mathcal{N}_p^{(A'_1)}(s) := \frac{1}{p^{2s} - \chi^{(A'_1)}(p)} \begin{pmatrix} p-1 & p^s - \chi^{(A'_1)}(p)p^{1-s} \\ p^s - \chi^{(A'_1)}(p)p^{1-s} & \chi^{(A'_1)}(p) \cdot (p-1) \end{pmatrix}.$$

Here  $\mathcal{N}_{p_{w+1}}^{(A'_1)}(s) \otimes \cdots \otimes \mathcal{N}_{p_v}^{(A'_1)}(s)$  denotes the Kronecker product of the matrices; see (2.48) below for the precise interpretation. We note that if  $\chi = 1$ , then  $\Phi(s) = B_{1,1}(s)$  and  $L(s, \chi^{(1)}) = \zeta(s)$ , and Lemma 2.9 specializes to the formula in [18, p. 536 (4.2)].

*Proof.* Given any  $A_1, A_2 \mid N$  we evaluate  $\varphi_{A_1 A_2}(s)$  by generalizing [18, pp. 534–535] to the case of an arbitrary character  $\chi$ ; we obtain, with  $B_j = N/A_j$ :

$$\begin{aligned} \varphi_{A_1 A_2}(s) &= \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \sum_{\substack{c_0 > 0 \\ (c_0, (B_1, A_2)(A_1, B_2)) = 1}} \sum_{\substack{0 \leq d_0 < (A_1, A_2)(B_1, B_2)c_0 \\ (d_0, (A_1, A_2)(B_1, B_2)c_0) = 1}} [\chi_{(A_1, A_2)} \overline{\chi_{(B_1, B_2)}}](d_0(B_1, A_2)) \\ &\quad \times \frac{[\chi_{(A_1, B_2)} \overline{\chi_{(A_2, B_1)}}](c_0(B_1, B_2))}{(c_0(B_1, B_2) \sqrt{A_1 A_2})^{2s}}. \end{aligned}$$

Write  $A'_j = (A_j, q(\chi))$  and  $A''_j = A_j/A'_j$ . The inner sum over  $d_0$  vanishes if  $\chi_{(A_1, A_2)} \overline{\chi_{(B_1, B_2)}} \neq 1$ , i.e. if  $A'_1 A'_2 \neq q(\chi)$ . In the remaining case,  $A'_1 A'_2 = q(\chi)$ , the full expression can be factored as follows, with  $B''_j = N/(q(\chi)A''_j)$ :

$$(2.48) \quad \begin{aligned} \varphi_{A_1 A_2}(s) &= \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{L(2s - 1, \chi^{(A'_1)})}{L(2s, \chi^{(A'_1)})} \cdot q(\chi)^{-s} \cdot \chi^{(A'_1)}((B''_1, B''_2)) \\ &\quad \times \prod_{p \mid (A''_1, A''_2)(B''_1, B''_2)} \frac{p-1}{p^{2s} - \chi^{(A'_1)}(p)} \prod_{p \mid (A''_1, B''_2)(A''_2, B''_1)} \frac{p^s - \chi^{(A'_1)}(p)p^{1-s}}{p^{2s} - \chi^{(A'_1)}(p)}. \end{aligned}$$

This agrees with the formula stated in the lemma.  $\square$

It follows from Lemma 2.9 and  $\chi(V^{-1}U_A) = \chi(T_{A_0}^{-1}) = (-1)^\varepsilon \chi_A(-1)$  (see (2.45)) that the matrix  $\Phi^\Gamma(s)$  from (2.33), (2.32) equals the block matrix  $(B_{A'_1, A'_2}(s))$  where now  $A'_1, A'_2$  run through the set  $\{d \mid q(\chi) : (-1)^\varepsilon \chi_d(-1) = 1\}$ . We note

$$\det \mathcal{N}_p^{(A'_1)}(s) := \frac{\chi^{(A'_1)}(p) p^{2-2s} - 1}{p^{2s} - \chi^{(A'_1)}(p)}.$$

We also remark, regarding the first factors in  $B_{A'_1, A'_2}(s)$ , that by the functional equation for the Dirichlet  $L$ -function we have

$$(2.49) \quad \sqrt{\pi} \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{L(2s - 1, \chi^{(A'_1)})}{L(2s, \chi^{(A'_1)})} \cdot q(\chi)^{-s} = g(\chi^{(A'_1)}) \frac{\Gamma(1-s) L(2-2s, \overline{\chi^{(A'_1)}})}{\Gamma(s) L(2s, \chi^{(A'_1)})} \cdot \pi^{2s-1} q(\chi)^{1-3s},$$

where  $g(\chi^{(A'_1)})$  is the Gauss sum  $g(\chi^{(A'_1)}) = \sum_{a \bmod q(\chi)} \chi^{(A'_1)}(a) \cdot e(a/q(\chi))$ . It follows that

$$\det B_{A'_1, A'_2}(s) = c \cdot \left( \frac{\Gamma(1-s) L(2-2s, \overline{\chi^{(A'_1)}})}{\Gamma(s) L(2s, \chi^{(A'_1)})} \cdot \pi^{2s} q(\chi)^{-3s} \right)^{2^{v-w}} \cdot \prod_{j=w+1}^v \left( \frac{\chi^{(A'_1)}(p_j) p_j^{2-2s} - 1}{p_j^{2s} - \chi^{(A'_1)}(p_j)} \right)^{2^{v-w-1}}.$$

for some constant  $c$  (viz.,  $c$  is independent of  $s$ ). Note

$$\frac{d}{ds} \log \left( \frac{\chi^{(A)}(p) p^{2-2s} - 1}{p^{2s} - \chi^{(A)}(p)} \right) = 2(\log p) \cdot \left( \frac{1}{\chi^{(A)}(p)^{-1} p^{2s-2} - 1} + \frac{1}{\chi^{(A)}(p) p^{-2s} - 1} \right).$$

Hence (using the symmetry relation  $\overline{\chi^{(A)}} = \chi^{(q(\chi)/A)}$ ):

$$\begin{aligned} \frac{(\varphi^\Gamma)'(s)}{\varphi^\Gamma(s)} &= \sum_{\substack{A|q(\chi) \\ \chi_A(-1)=(-1)^\varepsilon}} 2^{v-w} \cdot \left\{ -\frac{\Gamma'(1-s)}{\Gamma(1-s)} - \frac{\Gamma'(s)}{\Gamma(s)} - 2 \frac{L'(2-2s, \chi^{(A)})}{L(2-2s, \chi^{(A)})} - 2 \frac{L'(2s, \chi^{(A)})}{L(2s, \chi^{(A)})} \right. \\ &\quad \left. + 2 \log \pi - 3 \log q(\chi) + \sum_{j=w+1}^v (\log p_j) \cdot \left( \frac{1}{\chi^{(A)}(p_j) p_j^{2s-2} - 1} + \frac{1}{\chi^{(A)}(p_j) p_j^{-2s} - 1} \right) \right\}. \end{aligned}$$

Now  $\frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \cdot \frac{(\varphi^\Gamma)'(\frac{1}{2}+ir)}{\varphi^\Gamma(\frac{1}{2}+ir)} dr$  from (2.37) can be evaluated using the two formulas:

$$\begin{aligned} \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \cdot \left( \frac{L'(1-2ir, \chi^{(A)})}{L(1-2ir, \chi^{(A)})} + \frac{L'(1+2ir, \chi^{(A)})}{L(1+2ir, \chi^{(A)})} \right) dr \\ = \begin{cases} \frac{1}{4} h(0) & \text{if } \chi = 1 \\ 0 & \text{if } \chi \neq 1 \end{cases} - \sum_{n=1}^{\infty} \frac{\Lambda(n) \chi^{(A)}(n)}{n} g(2 \log n) \end{aligned}$$

(which follows by imitating [18, p. 509]), and

$$\frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) \left( \frac{1}{\chi^{(A)}(p) p^{-1+2ir} - 1} + \frac{1}{\chi^{(A)}(p) p^{-1-2ir} - 1} \right) dr = -g(0) - \sum_{m=1}^{\infty} \frac{\chi^{(A)}(p)}{p^m} g(2 \log p^m)$$

(which follows using  $\frac{1}{x-1} = -1 - \sum_{m=1}^{\infty} x^m$  for  $|x| < 1$ ).

Furthermore we note that  $\text{Tr } \Phi^\Gamma(\frac{1}{2}) = 0$  whenever  $\chi \neq 1$ , for then  $\Phi^\Gamma(s)$  is 0 along the diagonal by Lemma 2.9. The same thing holds if  $\chi = 1$  and  $\varepsilon = 1$ , since then  $R_{\Gamma, \chi} = \emptyset$ . Finally if  $\chi = 1$  and  $\varepsilon = 0$  then from Lemma 2.9 and (2.49)  $\Phi^\Gamma(\frac{1}{2})$  equals minus the  $2^v \times 2^v$ -identity matrix and thus  $\text{Tr } \Phi^\Gamma(\frac{1}{2}) = -2^v$ .

Combining all the results obtained so far, we see that the contribution from the last four lines in (2.37) translate exactly into the last three lines of (2.39).

We now consider lines 2 and 3 in (2.37). The method we use to enumerate the conjugacy classes appearing in these sums is well-known, cf., e.g., [14], [43], [29]. Here we will follow the setup in [29] fairly closely.

Recall  $\Gamma = \Gamma_0^\pm(N)$ . Given  $T \in \Gamma$  with determinant  $n \in \{1, -1\}$  and trace  $t \in \mathbb{Z}$  (only defined up to sign change) one shows that  $T \in M$  (i.e.  $T$  does not fix any cusp, cf. (2.18)) if

and only if  $\sqrt{t^2 - 4n} \notin \mathbb{Q}$ . Also:

$$(2.50) \quad \begin{cases} \text{If } t^2 - 4n > 0, t \neq 0: & T \text{ is hyp. or glide refl.}; N(T) = \frac{1}{4}(|t| + \sqrt{t^2 - 4n})^2 \\ \text{If } t^2 - 4n < 0: & T \text{ is elliptic, and } \cos \theta(T) = |t|/2. \end{cases}$$

In particular, since  $t^2 - 4n > 0$  and  $t = 0$  imply  $\sqrt{t^2 - 4n} = 2 \in \mathbb{Q}$ , it follows that  $M$  does not contain any reflections.

We will use the alternative format of the sums, (2.23), pointed out in Remark 2.3. In particular, recall the definitions of  $\bar{\Gamma}^+$ ,  $\bar{\Gamma}$ ,  $\bar{M} \subset \bar{G}$  given there. By what we have noted above, we have a disjoint union

$$(2.51) \quad \bar{M} = \bigsqcup_{n \in \{1, -1\}} \bigsqcup_{\substack{t \in \mathbb{Z} \\ \sqrt{t^2 - 4n} \notin \mathbb{Q}}} H_{n,t}, \quad \text{where } H_n := \{T \in \bar{\Gamma} \mid \det T = n\}.$$

Given any  $n, t$  as in (2.51), we let  $d, \ell$  be the unique integers such that  $t^2 - 4n = d\ell^2$ ,  $\ell > 0$ , and  $d$  is a *fundamental discriminant*, viz. either  $d \equiv 1 \pmod{4}$ ,  $d$  is squarefree and  $d \neq 1$ , or else  $d \equiv 0 \pmod{4}$ ,  $d/4$  is squarefree and  $d/4 \not\equiv 1 \pmod{4}$ . One shows easily that  $H_{n,t} \neq \emptyset$  if and only if  $\left[ p \mid \ell \text{ or } \left(\frac{d}{p}\right) \neq -1 \right]$  holds for each prime  $p \mid N$ . (This is a direct extension of [39, Lemma 3.4].) We now fix some  $n, t$  satisfying this condition, and fix some element  $T_t \in H_{n,t}$ . Using  $T_t^2 - tT_t + n = 0$  we see that the subalgebra  $\mathbb{Q}[T_t] \subset M_2(\mathbb{Q})$  is actually a quadratic field, isomorphic to  $\mathbb{Q}(\sqrt{d})$ . We write  $\omega := \frac{1}{2}(d + \sqrt{d})$ , so that each order in  $\mathbb{Q}(\sqrt{d})$  can be expressed as  $\mathfrak{r}[f] := \mathbb{Z} + f\omega\mathbb{Z}$  for some  $f \in \mathbb{Z}^+$ ; in particular  $\mathfrak{r}[1]$  is the ring of integers in  $\mathbb{Q}(\sqrt{d})$ . Using  $\mathbb{Q}[T_t] \cong \mathbb{Q}(\sqrt{d})$ , we write  $\mathfrak{r}[f]$  also for the corresponding order in  $\mathbb{Q}[T_t]$ . (This does not depend on which of the two possible isomorphisms  $\mathbb{Q}[T_t] \rightarrow \mathbb{Q}(\sqrt{d})$  we choose.)

Let  $R = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\}$ ; this is an order in the algebra  $M_2(\mathbb{Q})$  such that  $\bar{\Gamma}^+ = R \cap \text{SL}(2, \mathbb{Z}) = R^1$ . By [29, Lemma 5.2.2(2)] we have

$$C(T_t) := \{\delta T_t \delta^{-1} \mid \delta \in \text{GL}(2, \mathbb{Q})\} = \{A \in M_2(\mathbb{Q}) \mid \text{tr } A = t, \det A = n\}.$$

Note that for any  $\delta \in \text{GL}(2, \mathbb{Q})$ , the set  $\mathbb{Q}[T_t] \cap \delta^{-1}R\delta$  is an order in  $\mathbb{Q}[T_t]$ , and if  $\delta T_t \delta^{-1} = \delta' T_t \delta'^{-1}$  then  $\delta^{-1}\delta' \in \mathbb{Q}[T_t]$  by [29, Lemma 5.2.2(3)], and thus  $\mathbb{Q}[T_t] \cap \delta^{-1}R\delta = \mathbb{Q}[T_t] \cap \delta'^{-1}R\delta'$ . Hence  $C(T_t)$  can be decomposed as a disjoint union

$$(2.52) \quad C(T_t) = \bigsqcup_{f=1}^{\infty} C(T_t, \mathfrak{r}[f]),$$

where we have defined, for any order  $\mathfrak{r}$  in  $\mathbb{Q}[T_t]$ ,

$$(2.53) \quad C(T_t, \mathfrak{r}) := \{\delta T_t \delta^{-1} \mid \delta \in \text{GL}(2, \mathbb{Q}), \mathbb{Q}[T_t] \cap \delta^{-1}R\delta = \mathfrak{r}\}.$$

One also checks that if  $T_t \in \mathfrak{r}$  then  $C(T_t, \mathfrak{r}) \subset H_{n,t}$ , whereas  $T_t \notin \mathfrak{r}$  implies  $C(T_t, \mathfrak{r}) \cap H_{n,t} = \emptyset$ . Using  $T_t^2 - tT_t + n = 0$  we also see that  $T_t \in \mathfrak{r}[f]$  holds if and only if  $f \mid \ell$ . Hence

$$(2.54) \quad H_{n,t} = \bigsqcup_{f \mid \ell} C(T_t, \mathfrak{r}[f]).$$

Note that since  $\bar{\Gamma}^+ = R^1$ , each set  $C(T_t, \mathfrak{r})$  is closed under  $\bar{\Gamma}^+$ -conjugation.

The following is a direct generalization of [39, Lemma 3.7] (same proof):

**Lemma 2.10.** *In the situation above, take  $f \mid \ell$  and  $T \in C(T_t, \mathfrak{r}[f]) \subset H_{n,t}$ . If  $T$  is elliptic, i.e. if  $d < 0$ , then*

$$|Z_{\overline{\Gamma}^+}(T)| = |\mathfrak{r}[f]^1| = \frac{|\mathfrak{r}[1]^1|}{[\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1]}.$$

*If  $T$  is hyperbolic or a glide reflection, i.e. if  $d > 0$ , then for any  $T_0 \in \overline{M}$  with  $Z_{\overline{\Gamma}^+}(T) = \{\pm T_0^k \mid k \in \mathbb{Z}\}$  we have*

$$\log N(T_0) = 2[\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1] \cdot \log \epsilon_1,$$

*where  $\epsilon_1 > 1$  is the proper fundamental unit in  $\mathbb{Q}(\sqrt{d})$ .*

The  $\overline{\Gamma}^+$ -conjugacy classes in each  $C(T_t, \mathfrak{r})$  can be counted using a local-to-global principle: Let  $\mathfrak{r} = \mathfrak{r}[f]$  for some  $f \mid \ell$ . For each prime  $p$  we set

$$(2.55) \quad C_p(T_t, \mathfrak{r}) = \{xT_t x^{-1} \mid x \in \mathrm{GL}(2, \mathbb{Q}_p), \mathbb{Q}_p[T_t] \cap x^{-1}R_p x = \mathfrak{r}_p\}$$

where  $R_p = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in M_2(\mathbb{Z}_p) \mid \gamma \equiv 0 \pmod{N} \right\}$  is the closure of  $R$  in  $M_2(\mathbb{Q}_p)$  and  $\mathfrak{r}$  is the closure of  $\mathfrak{r}$  in  $\mathbb{Q}_p(\sqrt{d}) \cong \mathbb{Q}_p[T_t]$ . One checks that  $C_p(T_t, \mathfrak{r})$  is closed under  $R_p^\times$ -conjugation. We also set

$$C_\infty(T_t, \mathfrak{r}) = \{xT_t x^{-1} \mid x \in \mathrm{GL}(2, \mathbb{R})\},$$

which is closed under  $R_\infty^\times$ -conjugation, where we have set  $R_\infty^\times := \mathrm{GL}^+(2, \mathbb{R})$ . Let  $C(T_t, \mathfrak{r})//\overline{\Gamma}$  denote the set of  $\overline{\Gamma}^+$ -conjugacy classes in  $C(T_t, \mathfrak{r})$ , and define  $C_v(T_t, \mathfrak{r})//R_v^\times$  similarly for each place  $v$  of  $\mathbb{Q}$ . Clearly we then have a natural map

$$(2.56) \quad \theta : C(T_t, \mathfrak{r})//\overline{\Gamma}^+ \rightarrow \prod_{v \in \{\infty, 2, 3, 5, \dots\}} C_v(T_t, \mathfrak{r})//R_v^\times.$$

By [29, Lemma 6.5.2] (which is trivially generalized to also cover the cases of hyperbolic elements and glide reflections), the map  $\theta$  is surjective, and in fact  $\theta$  is exactly  $\mathbf{h}^+(\mathfrak{r})$ -to-1, where  $\mathbf{h}^+(\mathfrak{r})$  is the (narrow) class number for  $\mathfrak{r}$ . The case  $v = \infty$  is easy; as in [29, (6.6.1)] we have

$$|C_\infty(T_t, \mathfrak{r}[f])//R_\infty^\times| = \begin{cases} 2 & \text{if } d < 0 \\ 1 & \text{if } d > 0. \end{cases}$$

Also for primes  $p \nmid N$  we have  $R_p = M_2(\mathbb{Z}_p) \subset M_2(\mathbb{Q}_p)$  and thus  $|C_p(T_t, \mathfrak{r})//R_p^\times| = 1$  (cf., e.g., [29, Theorem 6.6.7]).

Finally, if  $p \mid N$ , a complete set of representatives for  $C_p(T_t, \mathfrak{r}[f])//R_p^\times$  is given in [29, Thm. 6.6.6]<sup>2</sup>. We should take  $\nu = 1$  therein (see [29, (6.6.3)]), since  $N$  is squarefree, and  $\rho = \mathrm{ord}_p(\ell/f)$ . Then notice that  $t^2 - 4n = \ell^2 d$  is divisible by  $p^{2\rho}$ . Making the result slightly more explicit in a similar way as in [39, (3.26) and below], we obtain that

$$(2.57) \quad |C_p(T_t, \mathfrak{r}[f])//R_p^\times| = \begin{cases} 2 & \text{if } p \mid f \\ 1 + \left(\frac{d}{p}\right) & \text{if } p \nmid f \end{cases}$$

and that if  $C_p(T_t, \mathfrak{r}[f]) \neq \emptyset$  (i.e.  $p \mid f$  or  $(d/p) \neq -1$ ) and  $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$  runs through any set of representatives for  $C_p(T_t, \mathfrak{r}[f])//R_p^\times$ , then  $\delta \pmod{p}$  runs through all the solutions of the

<sup>2</sup>There is a misprint in the statement of [29, Thm. 6.6.6]: “ $\Omega'/p^{\nu+\rho+1}$ ” should be replaced by  $\Omega'/p^{\nu+\rho}$ , i.e. a complete set of representatives of  $\Omega'(\alpha, \mathfrak{r}_p) \pmod{p^{\nu+\rho}}$ .

congruence equation  $\delta^2 - t\delta + n \equiv 0 \pmod{p}$ , visiting each solution the same number of times.

Using these observations and our remarks about the map  $\theta$  in (2.56) it now follows that

$$(2.58) \quad |C(T_t, \mathfrak{r}[f])//\overline{\Gamma}^+| = \mathbf{h}^+(\mathfrak{r}[f]) \cdot \left\{ \begin{array}{ll} 2 & \text{if } d < 0 \\ 1 & \text{if } d > 0 \end{array} \right\} \cdot \prod_{p|N} \left\{ \begin{array}{ll} 2 & \text{if } p \mid f \\ 1 + \left(\frac{d}{p}\right) & \text{if } p \nmid f \end{array} \right\},$$

and that if  $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$  runs through any set of representatives for  $C(T_t, \mathfrak{r}[f])//\overline{\Gamma}^+$ , then  $\delta \pmod{N}$  runs through all solutions to the congruence equation  $\delta^2 - t\delta + n \equiv 0 \pmod{N}$ , visiting each solution the same number of times.

Using (2.50), (2.51), (2.54), Lemma 2.10 and (2.58) and  $\chi\left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\right) = (\alpha\delta - \beta\gamma)^\varepsilon \chi(\delta)$ , we finally see that lines 2 and 3 in (2.37) can be expressed explicitly as in lines 2 to 3 of (2.39).

This concludes the proof of Theorem 3.

**2.3. Sieving for newforms.** From now on we keep  $\chi$  and  $\varepsilon \in \{0, 1\}$  fixed, with  $\chi$  an even Dirichlet character of squarefree conductor  $q(\chi)$ . This specifies a character on the group  $\Gamma_0^\pm(q(\chi))$ . We let  $N$  vary through arbitrary squarefree positive integers divisible by  $q(\chi)$ . Given  $\lambda \geq 0$  we let  $\mathbf{A}(N, \lambda)$  denote the linear space of  $(\Gamma_0^\pm(N), \chi)$ -Maass waveforms with eigenvalue  $\lambda$ . Let  $\delta_\ell$  be the natural map  $\mathbf{A}(N, \lambda) \rightarrow \mathbf{A}(N\ell, \lambda)$  given by  $\delta_\ell f(z) := f(\ell z)$ . Recall that the *space of newforms*,  $\mathbf{A}_{\text{new}}(N, \lambda)$ , is defined as the orthogonal complement in  $\mathbf{A}(N, \lambda)$  of the linear span of all spaces  $\delta_\ell \mathbf{A}(M, \lambda)$  with  $q(\chi) \mid M$ ,  $\ell M \mid N$ ,  $M < N$ .

In this section we apply a standard sieving procedure to obtain a version of the trace formula for  $\Gamma_0^\pm(N), \chi$  where we only sum over newform eigenvalues on the left hand side.

It is well-known that for each  $\lambda > 0$  and  $N$  divisible by  $q(\chi)$ ,  $\mathbf{A}(N, \lambda)$  can be expressed as a direct sum of subspaces as follows:

$$\mathbf{A}(N, \lambda) = \bigoplus_{M\ell \mid (N/q(\chi))} \delta_\ell \mathbf{A}_{\text{new}}(Mq(\chi), \lambda),$$

where the sum is taken over all pairs of positive integers  $M, \ell$  with  $M\ell$  dividing  $N/q(\chi)$ , cf. [28]. Hence

$$\dim \mathbf{A}(N_1 q(\chi), \lambda) = \sum_{M \mid N_1} d\left(\frac{N_1}{M}\right) \dim \mathbf{A}_{\text{new}}(Mq(\chi), \lambda), \quad \forall N_1 \in \mathbb{Z}^+, \lambda > 0.$$

Applying Möbius inversion to this formula we obtain

$$\dim \mathbf{A}_{\text{new}}(N_1 q(\chi), \lambda) = \sum_{M \mid N_1} \beta\left(\frac{N_1}{M}\right) \dim \mathbf{A}(Mq(\chi), \lambda), \quad \forall N_1 \in \mathbb{Z}^+, \lambda > 0,$$

where  $\beta(n)$  can be defined by  $\zeta(s)^{-2} = \sum_{n=1}^{\infty} \beta(n)n^{-s}$ ; note that for  $n$  squarefree we have  $\beta(n) = \mu(n)d(n) = (-2)^{\#\{p \mid n\}}$  where  $\#\{p \mid n\}$  denotes the number of prime factors of  $n$ . Hence to obtain the trace formula for  $(\Gamma_0^\pm(N_1 q(\chi)), \chi)$ -*newforms*, we should add the trace formulas for  $(\Gamma_0^\pm(Mq(\chi)), \chi)$ , weighted with  $\beta(N_1/M)$ , over all  $M \mid N_1$ . We keep  $N_1$  squarefree and relatively prime to  $q(\chi)$ , so that the final level  $N = N_1 q(\chi)$  is squarefree. Regarding the contribution from  $\lambda = 0$ , note that  $\lambda = 0$  occurs as a (simple) eigenvalue on

$(\Gamma_0^\pm(Mq(\chi)), \chi)$  if and only if  $\chi = 1$  and  $\varepsilon = 0$ , and then the contribution is

$$\sum_{M|N_1} \beta\left(\frac{N_1}{M}\right) h(-i/2) = \mu(N_1) h(-i/2).$$

The contributions from the other terms in (2.39) are evaluated in a fairly straightforward manner (similar to e.g. [39, §5]), and the final result is as follows.

**Theorem 4.** *If  $N = q(\chi)$  then all  $(\Gamma_0^\pm(N), \chi)$ -Maass waveforms are newforms, and hence the trace formula for  $(\Gamma_0^\pm(N), \chi)$ -newforms is given by (2.39). If  $N = N_1 q(\chi)$  with  $N_1 \geq 2$  squarefree and relatively prime to  $q(\chi)$ , then the trace formula for  $(\Gamma_0^\pm(N), \chi)$ -newforms is given by:*

$$\begin{aligned} (2.59) \quad & \left[ \text{If } \chi = 1, \varepsilon = 0: \mu(N_1) h(-i/2) \right] + \sum_{\lambda > 0} h(r_\lambda) \\ &= \frac{\prod_{p|q(\chi)} (p+1) \cdot \prod_{p|N_1} (p-1)}{24} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr \\ &+ \frac{1}{2} \sum_{n \in \{-1, 1\}} n^\varepsilon \sum_{\substack{t \in \mathbb{Z} \\ \sqrt{t^2 - 4n} \notin \mathbb{Q}}} \left\langle \chi(\delta) \right\rangle_{\substack{\delta^2 - t\delta + n \equiv 0 \\ \text{mod } q(\chi)}} \cdot \left( \sum_{\substack{f|\ell \\ (f, N_1) = 1}} \mathbf{h}^+(\mathfrak{r}[f]) \cdot [\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1] \right. \\ &\quad \cdot \prod_{p|q(\chi)} \left\{ \begin{array}{ll} 2 & \text{if } p \mid f \\ 1 + \left(\frac{d}{p}\right) & \text{if } p \nmid f \end{array} \right\} \prod_{p|N_1} \left( \left(\frac{d}{p}\right) - 1 \right) \cdot A(t, n) \\ &- d(q(\chi)) \cdot \left[ \text{If } N_1 \text{ prime: } \left\{ \begin{array}{ll} \frac{1}{2} + \frac{1}{2} C_{\chi, \varepsilon} & \text{if } N_1 > 2 \\ \frac{1}{4} + \frac{3}{4} C_{\chi, \varepsilon} & \text{if } N_1 = 2 \end{array} \right\} (\log N_1) g(0) \right. \\ &\quad \left. + \sum_{m=1}^{\infty} \frac{\log N_1}{N_1^m} \cdot \{\chi\}_\varepsilon(N_1^m) \cdot g(2m \log N_1) \right]. \end{aligned}$$

The notation in this formula is the same as in Theorem 3.

**2.4. Computational remarks.** In our applications of the trace formula (see Section 4), we choose test functions to localize on the spectral side of (2.39) and (2.59). Correspondingly, we encounter long sums on the geometric side. For such test functions, the bottleneck in terms of computation is the contribution from hyperbolic terms, that is the sum over  $t \in \mathbb{Z}$  with  $t^2 \pm 4 > 0$ , which involves the quadratic class numbers  $\mathbf{h}(t^2 \pm 4)$ . For numerical purposes, it is best to consider test functions  $g$  of compact support; (2.39) and (2.59) are then exact, in the sense that one can compute the geometric side to arbitrary precision with a fixed list of class numbers. Precisely, given a list of the numbers  $\mathbf{h}(d)$  for  $d = t^2 \pm 4 < e^X$ , one can evaluate the trace formula for any  $g$  as in Theorem 2 with support contained in  $[-X, X]$ .

There are many algorithms for computing quadratic class numbers. In our implementation, we used a simple variant of the algorithm of [7], which has asymptotic complexity  $O(d^{1/4+\varepsilon})$  for computing  $\mathbf{h}(d)$ . While heuristically faster methods exist, they rely on GRH to certify the results; we avoided these so as to be able to state Theorem 1 unconditionally. However, note that no matter which class number algorithm is used, one cannot escape the exponential complexity of computing the hyperbolic sum; it is therefore unlikely that

the faster conditional methods would significantly extend the results of Theorem 1 without employing a new idea.

We computed  $\mathbf{h}(d)$  for all  $d = t^2 \pm 4 < e^{36}$ , using approximately two weeks of time on a network of 80 Linux PCs. Note that (2.39) and (2.59) are very sensitive to errors in the class number computations, since each of the approximately  $e^{18}$  values of  $\mathbf{h}(d)$  has roughly the same weight, yet they cancel out to yield an answer of small size. Thus, the trace formula itself serves as a good check of the correctness of the computation. (In fact, this observation may be extended to a fast heuristic algorithm for certifying a list of class numbers without relying on GRH; this theme is being explored in the case of imaginary quadratic fields by Jacobson and Ramachandran.) For the same reason, one must be careful to use sufficient precision to handle the catastrophic cancellation that occurs when evaluating the trace formula; for our implementation, we used the PC “long double” type, which has a 64-bit mantissa.

It is worthwhile to note that all terms of (2.39) and (2.59) may be expressed in terms of  $g$ . In particular,

$$\begin{aligned} h(0) &= \int_{-\infty}^{\infty} g(u) du; & \int_{-\infty}^{\infty} rh(r) \tanh(\pi r) dr &= - \int_{-\infty}^{\infty} \frac{g'(u)}{\sinh(u/2)} du; \\ A(0, 1) &= \frac{1}{2|\mathfrak{r}[1]^1|} \int_{-\infty}^{\infty} \frac{g(u)}{\cosh(u/2)} du; & A(1, 1) &= \frac{1}{|\mathfrak{r}[1]^1|} \int_{-\infty}^{\infty} g(u) \frac{\cosh(u/2)}{\cosh(u) + \frac{1}{2}} du; \\ \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'(1+ir)}{\Gamma(1+ir)} dr &= -\gamma g(0) + \int_0^{\infty} \log(u) g'(u) du \\ &+ \frac{1}{4} \int_{-\infty}^{\infty} g(u) du + \int_0^{\infty} \log\left(\frac{\sinh(u/2)}{u/2}\right) g'(u) du; \\ \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'(\frac{1}{2}+ir)}{\Gamma(\frac{1}{2}+ir)} dr &= -\gamma g(0) + \int_0^{\infty} \log(u) g'(u) du \\ &- \int_0^{\infty} \log\left(\frac{\sinh(u/2)}{u/2}\right) g'(u) du + 2 \int_0^{\infty} \log\left(\frac{\sinh(u/4)}{u/4}\right) g'(u) du. \end{aligned}$$

Cf. [17, pp. 27–28, 450], [18, pp. 205–206] and our references in connection with (2.27).

Since  $g$  has compact support, the integrals above may be computed accurately, either symbolically or by standard methods such as Gaussian quadrature. Moreover, for any fixed family of test functions, such as those of Section 4, the numerical integrations need only be performed once, and may then be used for any  $(N, \chi)$ .

Next, we note that the hyperbolic and parabolic terms of (2.39) and (2.59) have multiplicative forms that are faster to evaluate. Firstly, the class number  $\mathbf{h}^+(\mathfrak{r}[f])$  is related to the usual narrow class number  $\mathbf{h}^+(d) = \mathbf{h}^+(\mathfrak{r}[1])$ , where  $d$  is the fundamental discriminant, by  $\mathbf{h}^+(\mathfrak{r}[f]) \cdot [\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1] = \mathbf{h}^+(d) \cdot f \prod_{p|f} (1 - (\frac{d}{p})p^{-1})$ . Using this, one finds by a short computation that the sum over  $f \mid \ell$  (in either formula) is non-zero precisely when  $q_1 \mid q(\chi) \mid (t^2 - 4n)q_1$ , where  $q_1 = \prod_{p|N, (\frac{d}{p})=1} p$ , and in this case:

$$\sum_{\substack{f|\ell \\ (f, N_1)=1}} \mathbf{h}^+(\mathfrak{r}[f]) \cdot [\mathfrak{r}[1]^1 : \mathfrak{r}[f]^1] \prod_{p|q(\chi)} \left\{ \begin{array}{ll} 2 & \text{if } p \mid f \\ 1 + \left(\frac{d}{p}\right) & \text{if } p \nmid f \end{array} \right\} \prod_{p|N_1} \left( \left(\frac{d}{p}\right) - 1 \right)$$



$$\begin{aligned}
 &= d \left( \frac{N}{(N, d)} \right) \mu(N_1) \mathbf{h}^+(d) \prod_{\substack{p|\ell \\ p \nmid N_1}} \left( \begin{cases} 0 & \text{if } p \mid q(\chi) \text{ and } \left(\frac{d}{p}\right) = -1 \\ 1 & \text{else} \end{cases} \right) \\
 &\quad + \left( p - \left(\frac{d}{p}\right) \right) \frac{p^{\text{ord}_p(\ell)} - 1}{p - 1} \left( \begin{cases} 2 & \text{if } p \mid (q(\chi), d) \\ 1 & \text{else} \end{cases} \right).
 \end{aligned}$$

Since this expression depends only on  $q(\chi)$  and not  $\chi$  itself, it is more efficient to work out the trace formula for all  $\chi$  of a given conductor simultaneously. The dependence on  $\chi$  enters through  $\langle \chi(\delta) \rangle$  and  $\{\chi\}_\varepsilon$ , which also have multiplicative forms:

$$(2.60) \quad \langle \chi(\delta) \rangle_{\substack{\delta^2 - t\delta + n \equiv 0 \\ \text{mod } q(\chi)}} = \prod_{p|q(\chi)} \begin{cases} \text{Re } \chi_p(\delta_p) & \text{if } \chi_p \text{ is even or } n = 1, \\ \text{Im } \chi_p(\delta_p) & \text{if } \chi_p \text{ is odd and } n = -1, \end{cases}$$

where  $\delta_p$  is any root of  $\delta^2 - t\delta + n \pmod{p}$ , and

$$(2.61) \quad \frac{1}{2}(\{\chi\}_0 + \{\chi\}_1) = \prod_{p|q(\chi)} \text{Re } \chi_p; \quad \frac{1}{2}(\{\chi\}_0 - \{\chi\}_1) = \prod_{p|q(\chi)} \begin{cases} \text{Re } \chi_p & \text{if } \chi_p \text{ is even,} \\ \text{Im } \chi_p & \text{if } \chi_p \text{ is odd.} \end{cases}$$

Finally, we remark that the expressions are somewhat simpler if we compute the sum and difference of the formulas for  $\varepsilon = 0$  and  $\varepsilon = 1$  rather than computing them individually. In particular, there are only hyperbolic terms for one value of  $n \in \{-1, 1\}$ , and the parabolic terms simplify thanks to (2.61).

### 3. CM FORMS

By a well-known construction of Maass [26], given a Hecke character over a real quadratic field, one may associate a Maass form or Eisenstein series. In this section we classify the Maass forms arising in this manner, and determine their contribution to the trace formula. First, we recall some facts about Hecke characters.

**3.1. Hecke characters.** Let  $K = \mathbb{Q}(\sqrt{D})$  be a real quadratic field of fundamental discriminant  $D$  and fundamental unit  $\epsilon_0$ ,  $\mathcal{O}$  its ring of integers, and  $\mathfrak{m}$  an ideal of  $\mathcal{O}$ . We imbed  $K$  in  $\mathbb{R}^2$  as usual. Let  $J^{\mathfrak{m}}$  be the group of fractional ideals with numerator and denominator relatively prime to  $\mathfrak{m}$ . A Hecke character over  $K$  is a homomorphism  $\psi : J^{\mathfrak{m}} \rightarrow S^1$  such that there exist characters  $\psi_f : (\mathcal{O}/\mathfrak{m})^\times \rightarrow S^1$  and  $\psi_\infty : (\mathbb{R}^\times)^2 \rightarrow S^1$  such that  $\psi(a\mathcal{O}) = \psi_f(a)\psi_\infty(a)$  for all  $a \in \mathcal{O}$  relatively prime to  $\mathfrak{m}$ .

Note that given  $\psi_f$  and  $\psi_\infty$ , in order for there to be a Hecke character with these data,  $\psi_f(a)\psi_\infty(a)$  must define a character of principal ideals, i.e.

$$(3.1) \quad \psi_f(\epsilon)\psi_\infty(\epsilon) = 1 \text{ for all } \epsilon \in \mathcal{O}^\times.$$

We record the following facts, whose proofs may be found, for example, in Neukirch [30].

- For any Hecke character  $\psi$ , the associated  $\psi_f$  and  $\psi_\infty$  are unique.
- Any  $\psi_f$  and  $\psi_\infty$  satisfying (3.1) come from a Hecke character.
- Any two Hecke characters with the same  $\psi_f$  and  $\psi_\infty$  differ by a character of the ideal class group.

Since we are not concerned with calculating Hecke eigenvalues, it turns out that the basic data (level, nebentypus character, eigenvalue and parity) of the associated Maass forms may be computed directly from the character  $\psi_f$ , with no class group computations needed. The above points show that each set of data occurs with multiplicity  $\mathbf{h}(D)$ , the class number. However, it is possible for two distinct Hecke characters to induce the same Maass form; we return to this point in Section 3.2 below.

Given a character  $\psi_f$  we look for a matching  $\psi_\infty$  satisfying (3.1). In this real quadratic case,  $\psi_\infty$  always takes the form (cf. [30])

$$(3.2) \quad \psi_\infty(x_1, x_2) = (\operatorname{sgn} x_1)^{a_1} (\operatorname{sgn} x_2)^{a_2} |x_1|^{ir_1} |x_2|^{ir_2},$$

for certain  $a_1, a_2 \in \{0, 1\}$  and  $r_1, r_2 \in \mathbb{R}$ . The ones that give rise to weight 0 Maass forms have  $a_1 = a_2 = a$  and  $r_1 = -r_2 = r$ ; these numbers determine the parity and eigenvalue of each associated form  $f$ , respectively. If  $\psi_f(\epsilon_0) = e(\alpha)$ , the possible solutions to (3.1) are:

$$(3.3) \quad \begin{array}{|c|c|c|} \hline & a = 0 \text{ (even forms)} & a = 1 \text{ (odd forms)} \\ \hline N(\epsilon_0) = 1 & r = \frac{\pi}{\log \epsilon_0}(-\alpha + n) & r = \frac{\pi}{\log \epsilon_0}(-\alpha + n) \\ \hline N(\epsilon_0) = -1 & r = \frac{\pi}{\log \epsilon_0}(-\alpha + n) & r = \frac{\pi}{\log \epsilon_0}(-\alpha + \frac{1}{2} + n) \\ \hline \end{array}$$

where  $n$  is an arbitrary integer.

The level of the associated forms is  $DN(\mathfrak{m})$ . To determine the nebentypus character  $\chi$ , we consider the  $L$ -function  $L(s, f)$ , where  $\chi(p)$  appears as the coefficient of  $p^{-2s}$  in the Euler product. For split primes  $p\mathcal{O} = \mathfrak{p}_1\mathfrak{p}_2$ , we have  $\chi(p) = \psi(\mathfrak{p}_1)\psi(\mathfrak{p}_2) = \psi(p\mathcal{O}) = \psi_f(p)$ , while for inert primes  $p$ ,  $\chi(p) = -\psi(p\mathcal{O}) = -\psi_f(p)$ . In other words,

$$(3.4) \quad \chi = \psi_f|_{\mathbb{Z}/N(\mathfrak{m})\mathbb{Z}} \cdot \chi_D,$$

where  $\chi_D$  is the quadratic character.

**3.2. Coincident forms and cuspidality.** As mentioned above, it is possible for two distinct Hecke characters to induce the same Maass form. One way is as follows.

Let  $S$  be the set of Hecke characters of any modulus  $\mathfrak{m}$  with  $N(\mathfrak{m}) = N/D$ . If  $\tau$  is the non-trivial automorphism of  $K$ , then  $\tau$  acts on  $S$  by  $\psi \mapsto \psi^\tau : \psi^\tau(\mathfrak{a}) = \psi(\mathfrak{a}^\tau)$ . (Note that if  $\psi$  has conductor  $\mathfrak{m}$  then  $\psi^\tau$  has conductor  $\mathfrak{m}^\tau$ .) Since  $\psi(p\mathcal{O}) = \psi^\tau(p\mathcal{O})$  for all inert primes, and  $\psi(\mathfrak{p}) + \psi(\mathfrak{p}^\tau)$  is the same with  $\psi$  replaced by  $\psi^\tau$ , we see that  $\psi$  and  $\psi^\tau$  have the same  $L$ -function, and therefore induce the same Maass form (or Eisenstein series).

Essentially, that is the only way for this coincidence to occur. To see that, let  $\psi_1$  and  $\psi_2$  be Hecke characters with the same  $L$ -function. In particular, their  $r$ -values must agree up to sign, and replacing  $\psi_2$  by  $\psi_2^\tau$  if necessary we may assume that they are equal. For  $r = 0$  the situation is in general complicated, as there may be Hecke characters over *different fields* with the same  $L$ -function. In what follows, we treat this case separately by counting distinct ray class character  $L$ -functions with PARI, cf. [8, rayclass.gp]. (Strictly speaking, we only produce a *lower bound* for the number of these forms associated to dihedral Galois representations, since we compare the Satake parameters of the  $L$ -functions at a short list of primes. In principle it would be possible to find an effective set of primes that need to be checked in order to conclude that our list is complete. However, that is unnecessary, as we obtain an upper bound from the trace formula in the process of verifying Selberg's conjecture.) For  $r \neq 0$ , equality of the  $r$ -values implies by (3.3) that the ratio of regulators of the two fields is a rational number, which can only happen if they are the same field.

Next we consider the Euler factors of the  $L$ -functions at each rational prime  $p$ . When  $p$  is inert, we have  $\psi_1(p\mathcal{O}) = \psi_2(p\mathcal{O})$ . For  $p\mathcal{O} = \mathfrak{p}\mathfrak{p}^\tau$  split, we have equality of the sets of Satake parameters:  $\{\psi_1(\mathfrak{p}), \psi_1(\mathfrak{p}^\tau)\} = \{\psi_2(\mathfrak{p}), \psi_2(\mathfrak{p}^\tau)\}$ . In all cases, we see that either  $\psi_2(\mathfrak{p}) = \psi_1(\mathfrak{p})$  or  $\psi_2(\mathfrak{p}) = \psi_1^\tau(\mathfrak{p})$ . If  $\psi_1 = \psi_1^\tau$  then we are done. Otherwise, if  $\psi_2$  is distinct from both  $\psi_1$  and  $\psi_1^\tau$ , then it is easy to see using  $L$ -function techniques that these conditions must fail simultaneously for infinitely many  $\mathfrak{p}$ . Briefly, we consider the sum

$$(3.5) \quad \sum_{\substack{N(\mathfrak{p}) \leq X \\ \mathfrak{p} \nmid \mathfrak{m}\mathfrak{m}^\tau}} (\psi_1^\tau(\mathfrak{p}) - \psi_2(\mathfrak{p})) \overline{(\psi_1(\mathfrak{p}) - \psi_2(\mathfrak{p}))} \log N(\mathfrak{p}).$$

The product expands to  $1 + \psi_1^\tau \overline{\psi_1} - \psi_1^\tau \overline{\psi_2} - \overline{\psi_1} \psi_2$ , and each of these last three characters is non-trivial by hypothesis. Thus, by the prime number theorem for Hecke characters, we see that (3.5) is asymptotic to  $X$ .

Finally, we recall that the induced form  $f$  is cuspidal if and only if  $\psi \neq \psi^\tau$ . If  $\psi = \psi^\tau$  then we have  $r = 0$  and  $\mathfrak{m} = \mathcal{O}$ , i.e.  $\psi$  is a character of the narrow class group. Moreover, in this case  $\psi^2(\mathfrak{a}) = \psi(\mathfrak{a}\mathfrak{a}^\tau) = 1$ . Thus, the non-cuspidal forms correspond precisely to the 2-torsion narrow class group characters. We eliminate those elements in our PARI program [8, rayclass.gp].

**3.3. Contribution of CM forms to the trace formula.** Now starting with a given (squarefree) level  $N$  and nebentypus character  $\chi$ , we determine the contribution to the trace formula from infinite order Hecke characters.

First, we must consider all fundamental discriminants  $D$  dividing  $N$ ; since  $N$  is squarefree, this means  $D \equiv 1 \pmod{4}$ . In order for there to exist forms with nebentypus  $\chi$ , we must have the factorization (3.4), i.e.  $\chi \cdot \chi_D$  must be a character of conductor exactly  $N/D$ . This then determines  $\psi_f|_{\mathbb{Z}/N(\mathfrak{m})\mathbb{Z}}$ . To pass to the character  $\psi_f$  we consider all ideals  $\mathfrak{m}$  of norm  $N/D$ . Since  $N$  is squarefree,  $N/D$  must be a product of split primes, i.e.  $p$  such that  $\chi_D(p) = 1$ .

Now for a given  $\mathfrak{m}$  of norm  $N/D$  we have  $\mathcal{O}/\mathfrak{m} \cong \mathbb{Z}/(N/D)\mathbb{Z}$ , so that  $\chi \cdot \chi_D$  lifts to a unique character  $\psi_f$  modulo  $\mathfrak{m}$ . Once defined, we have seen that the Hecke characters associated to  $\psi_f$  induce  $\mathbf{h}(D)$  Maass forms (ignoring possible coincidence of forms) with parameter  $r$  given by (3.3), where  $\psi_f(\epsilon_0) = e(\alpha)$ . We calculate the contribution of these forms to the trace formula, removing those of eigenvalue  $\frac{1}{4}$ , by Poisson sum:

$$(3.6) \quad \begin{aligned} \mathbf{h}(D) \sum_{\substack{n \in \mathbb{Z} \\ n \neq \alpha}} h\left(\frac{\pi}{\log \epsilon_0}(-\alpha + n)\right) \\ = 2\mathbf{h}(D) \log \epsilon_0 \sum_{n \in \mathbb{Z}} \psi_f(\epsilon_0^n) g(2 \log \epsilon_0 \cdot n) - \left[ \text{If } \psi_f(\epsilon_0) = 1: \quad \mathbf{h}(D)h(0) \right] \end{aligned}$$

and

$$(3.7) \quad \begin{aligned} \mathbf{h}(D) \sum_{\substack{n \in \mathbb{Z} \\ n \neq \alpha - 1/2}} h\left(\frac{\pi}{\log \epsilon_0}(-\alpha + 1/2 + n)\right) \\ = 2\mathbf{h}(D) \log \epsilon_0 \sum_{n \in \mathbb{Z}} (-1)^n \psi_f(\epsilon_0^n) g(2 \log \epsilon_0 \cdot n) - \left[ \text{If } \psi_f(\epsilon_0) = -1: \quad \mathbf{h}(D)h(0) \right]. \end{aligned}$$

Next we consider the sum of (3.6) and (3.7) over all possible lifts  $\psi_f$ . If  $N/D = p_1 \cdots p_s$ , then by the Chinese remainder theorem we may write the character  $\psi_f$  as a product  $\psi_f =$

$\psi_1 \cdots \psi_s$ , where  $\psi_i$  is a character of  $(\mathcal{O}/\mathfrak{p}_i)^\times \cong \mathbb{F}_{p_i}^*$ , and  $\mathfrak{m} = \prod \mathfrak{p}_i$ . To choose  $\mathfrak{p}_i$  amounts to picking a square root of  $D$  modulo  $p_i$ . If  $N(\epsilon_0) = 1$ , the different choices of  $\mathfrak{p}_i$  give rise to conjugate values of  $\psi_i(\epsilon_0)$ . Hence, we get

$$(3.8) \quad \sum_{\psi_f: \psi_f|_{\mathbb{Z}/(N/D)\mathbb{Z}} = \chi \cdot \chi_D} \psi_f(\epsilon_0^n) = d(N/D) \prod_{i=1}^s \operatorname{Re} \psi_i(\epsilon_0^n).$$

If  $N(\epsilon_0) = -1$ , we have similarly

$$(3.9) \quad \sum_{\psi_f: \psi_f|_{\mathbb{Z}/(N/D)\mathbb{Z}} = \chi \cdot \chi_D} \psi_f(\epsilon_0^n) = d(N/D) \prod_{i=1}^s \begin{cases} \operatorname{Re} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = 1, \\ \operatorname{Im} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = -1. \end{cases}$$

(Compare these expressions to (2.60) and (2.61).)

We substitute these into (3.6) and (3.7), and divide by 2 to correct for the coincidence of forms observed in Section 3.2. If  $N(\epsilon_0) = 1$ , for both even and odd forms we have

$$d\left(\frac{N}{D}\right) \mathbf{h}(D) \log \epsilon_0 \left( g(0) + 2 \sum_{n=1}^{\infty} g(2 \log \epsilon_0 \cdot n) \prod_{i=1}^s \operatorname{Re} \psi_i(\epsilon_0^n) \right) - \frac{\mathbf{h}(D)}{2} \#\{\psi_f | \psi_f(\epsilon_0) = 1\} h(0).$$

If  $N(\epsilon_0) = -1$ , we have for even forms

$$d\left(\frac{N}{D}\right) \mathbf{h}(D) \log \epsilon_0 \left( g(0) + 2 \sum_{n=1}^{\infty} g(2 \log \epsilon_0 \cdot n) \prod_{i=1}^s \begin{cases} \operatorname{Re} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = 1 \\ \operatorname{Im} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = -1 \end{cases} \right) - \frac{\mathbf{h}(D)}{2} \#\{\psi_f | \psi_f(\epsilon_0) = 1\} h(0),$$

and for odd forms

$$d\left(\frac{N}{D}\right) \mathbf{h}(D) \log \epsilon_0 \left( g(0) + 2 \sum_{n=1}^{\infty} (-1)^n g(2 \log \epsilon_0 \cdot n) \prod_{i=1}^s \begin{cases} \operatorname{Re} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = 1 \\ \operatorname{Im} \psi_i(\epsilon_0^n) & \text{if } \psi_i(-1)^n = -1 \end{cases} \right) - \frac{\mathbf{h}(D)}{2} \#\{\psi_f | \psi_f(\epsilon_0) = -1\} h(0).$$

**3.4. Multiplicity of non-CM forms.** The CM forms may be characterized in another way: they are the forms that admit a self-twist, i.e.  $f$  such that  $f = f \times \sigma$  for some (quadratic) Dirichlet character  $\sigma$ . This follows from the cuspidality criterion for the symmetric square lift; see [15, Theorem 9.3 and Remark 9.9]. This property makes them the fixed points of certain maps that we recall here.

Let  $N$  be squarefree, as above, and  $\chi$  a Dirichlet character mod  $N$ . Given any divisor  $A | N$  we define

$$(3.10) \quad \chi^{[A]} = \chi_A \cdot \overline{\chi_{N/A}}$$

(note that this is the Dirichlet character mod  $N$  which comes from the primitive Dirichlet character  $\chi^{(A)}$  defined on p. 14).

We now recall from [1], [2] (cf. also [35, §1.2.8] for the non-holomorphic case) the definition of the Atkin-Lehner operators. For  $Q \in \mathbb{Z}^+$  with  $Q | N$  we set

$$W_Q = \begin{pmatrix} Qx & y \\ Nz & Qw \end{pmatrix},$$

where  $x, y, z, w \in \mathbb{Z}$ ,  $y \equiv 1 \pmod{Q}$ ,  $x \equiv 1 \pmod{N/Q}$  and  $Q^2xw - Nzy = Q$ . Define  $W_Q f := f \circ W_Q$ . Then  $W_Q$  gives a bijection from the space  $\mathbf{A}_{\text{new}}(N, \chi, \lambda)$  of newforms of level  $N$ , eigenvalue  $\lambda$  and nebentypus character  $\chi$  to the corresponding space  $\mathbf{A}_{\text{new}}(N, \overline{\chi^{[Q]}})$ . (The inverse map is  $\chi_Q(-1)\chi_{N/Q}(Q) \cdot W_Q$ .) This map does not depend on the choice of  $x, y, z, w$ . Furthermore, for any prime  $q$ , if  $f \in \mathbf{A}_{\text{new}}(N, \chi, \lambda)$  satisfies  $T_q f = \lambda_q f$ , then  $T_q(W_Q f) = \overline{\chi_Q(q)} \cdot \lambda_q \cdot (W_Q f)$ . Also, by [2, Prop. 1.1] applied to  $W'_Q = \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} W_Q \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}$ , the parity of  $W_Q f$  is the same as that of  $f$  if and only if  $\chi_Q$  is even. In representation-theoretic terms, the map  $W_Q$  corresponds to twisting (the representation associated to)  $f$  by  $\overline{\chi_Q}$ .

If  $\chi$  is distinct from  $\overline{\chi^{[Q]}}$ , then we already separate the forms  $f$  and  $W_Q f$  in the trace formula. Therefore, the above map is most interesting when  $Q$  is divisible only by primes  $p$  for which  $\chi_p$  is quadratic, for then  $\chi$  and  $\overline{\chi^{[Q]}}$  agree. As mentioned above, the map can then have CM fixed points, but we remove the contribution of those from the trace formula.

For non-CM forms  $f$ , we have  $f \times \chi_1 \neq f \times \chi_2$  for distinct Dirichlet characters  $\chi_1$  and  $\chi_2$ . Therefore, the multiplicity of each non-CM eigenvalue is divisible by  $2^{\#\{p|N:\chi_p \text{ has order } 2\}}$ . If at least one of the  $\chi_p$  is odd quadratic, this multiplicity is spread equally over even and odd forms. Thus, to remove the multiplicity, after subtracting the contribution of CM forms we divide each trace formula by

$$\begin{cases} 2^{\#\{p|N:\chi_p \text{ has order } 2\}-1} & \text{if some } \chi_p \text{ is odd quadratic,} \\ 2^{\#\{p|N:\chi_p \text{ has order } 2\}} & \text{otherwise.} \end{cases}$$

#### 4. OPTIMIZATION

As discussed in Section 2.4, given a list of class numbers  $h(d)$  for  $d = t^2 \pm 4 \leq e^X$ , the trace formula gives a method to evaluate  $\sum_n h(r_n)$  for essentially any function  $h$  whose Fourier transform is supported in  $[-X, X]$ . When  $X$  is large, we may choose  $h$  to be narrowly concentrated around any particular point, and thus resolve features of the spectrum in places where the density of  $r$ -values is not too large compared to  $X$ . For a fixed support  $[-X, X]$ , there is a canonical way of choosing a “best” test function, which we describe in this section.

For  $r \in \mathbb{R} \cup i[-\frac{1}{2}, \frac{1}{2}]$ , let  $\mathcal{C}(X, r)$  be the set of trace-class functions  $h$ , with the corresponding  $g$  supported in  $[-X, X]$ , and the additional restrictions  $h(t) \geq 0$  for  $t \in \mathbb{R} \cup i[-\frac{1}{2}, \frac{1}{2}]$  and  $h(r) = 1$ . Define

$$(4.1) \quad F_X(r) := \inf_{h \in \mathcal{C}(X, r)} \sum_{n=1}^{\infty} h(r_n).$$

Note that if  $F_X(r) < 1$  for any value of  $X$  then  $\lambda = \frac{1}{4} + r^2$  cannot be an eigenvalue. Moreover, as  $X \rightarrow \infty$ ,  $F_X$  tends pointwise to the characteristic function of the eigenvalues. Thus, by evaluating  $F_X$  we can find provable intervals in which the eigenvalues must lie.

Although the definition of  $F_X$  is abstract, it is easy to construct concrete families of functions that closely approximate any desired function. For instance, let  $M$  be a large integer,  $\delta = X/2M$  and set

$$(4.2) \quad f(t) = \left( \frac{\sin \delta t/2}{\delta t/2} \right)^2 \sum_{n=0}^{M-1} x_n \cos \delta n t \quad \text{and} \quad h(t) = f(t)^2,$$

for arbitrary real numbers  $x_n$ . On the other side of the Fourier transform, this corresponds to taking  $g = \hat{f} * \hat{f}$ , where  $\hat{f}$  (the Fourier transform of  $f$ ) linearly interpolates arbitrary values at multiples of  $\delta$ .

The sum over eigenvalues in (4.1) is then a positive definite<sup>3</sup> quadratic form in the numbers  $x_n$ . To compute the matrix of the form, it suffices to compute the trace for the test functions  $h(t) = \left(\frac{\sin \delta t/2}{\delta t/2}\right)^4 \cos(\delta n t)$ ,  $0 \leq n \leq 2M - 2$ . The corresponding functions  $g$  are translates of a fixed function of small compact support. Since we may compute the formula for all localized test functions simultaneously, this computation takes roughly the same time as a single evaluation of the trace formula. Once the matrix is known, the infimum in (4.1) over this restricted class of test functions is easily found as the minimum of the quadratic form subject to the linear constraint  $f(r) = 1$ . This involves inverting the matrix, after which the minimum may be found quickly for many different values of  $r$ .

**4.1. Estimating low eigenvalues.** In this section we discuss a numerical method for computing eigenvalues using the trace formula. This problem was previously explored by Matthies, Steiner, [27] and Aurich, Steiner, [3]. Our technique, based on the observations presented above, has the advantage of giving rigorously provable estimates.

Our approach uses two different heuristics, which turn out to give the same estimates for the eigenvalues. First, we introduce some additional notation to be used in this section. Let  $V_r$  be the set of  $h = f^2$  as in (4.2), with the constraint  $f(r) = 1$ . Put  $y_n(t) = \left(\frac{\sin \delta t/2}{\delta t/2}\right)^2 \cos \delta n t$  and  $\mathbf{y}(t) = (y_0(t), \dots, y_{M-1}(t))$ . For  $\mathbf{x} = (x_0, \dots, x_{M-1})$ , let  $Q(\mathbf{x})$  denote the trace  $\sum_n h(r_n)$ , where  $h(t) = (\mathbf{x} \cdot \mathbf{y}(t))^2$  is the function corresponding to  $\mathbf{x}$ . Since  $Q$  is positive definite, there is a unique function  $h_r = f_r^2 \in V_r$  of minimal trace. Define  $q(r) = (\sum_n h_r(r_n))^{-1}$ .

As noted above,  $F_X$  tends to the characteristic function of the eigenvalues as  $X \rightarrow \infty$ . The first heuristic is that the eigenvalues should be near the peaks of  $F_X$  for large  $X$ . Our task is thus to locate the minima of  $q$ . While this could be accomplished by computing many sample points as described above, a more numerically stable method is to evaluate the derivative  $q'(r)$ , which is possible thanks to the following lemma.

**Lemma 4.1.**

$$(4.3) \quad \frac{d}{dr} \log q(r) = h'_r(r).$$

*Proof.* Let  $A$  be the symmetric matrix such that  $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ . Let  $\mathbf{x}(r)$  denote the coefficient vector corresponding to the optimal function  $h_r$ . By the method of Lagrange multipliers, we find that

$$(4.4) \quad q(r) = \mathbf{y}(r)^T A^{-1} \mathbf{y}(r) \quad \text{and} \quad \mathbf{x}(r) = A^{-1} \frac{\mathbf{y}(r)}{q(r)}.$$

In particular, note that  $q$  is a quadratic form in the dual variable  $\mathbf{y}$ . Differentiating, we have

$$(4.5) \quad q'(r) = 2\mathbf{y}'(r)^T A^{-1} \mathbf{y}(r) = 2q(r)\mathbf{y}'(r) \cdot \mathbf{x}(r) = 2q(r)f'_r(r) = q(r)h'_r(r). \quad \square$$

As (4.4) shows, the optimal functions  $h_r$  may be determined as easily as the value of  $q(r)$ . We may thus estimate the eigenvalues by searching for the negative-to-positive transitions

<sup>3</sup>As  $T \rightarrow \infty$ ,  $f(t)$  has  $O(T)$  zeros in  $[-T, T]$ , whereas  $\#\{r_n \leq T\} \gg T^2$ , by Weyl's law.

of the function  $h'_r(r)$ , and further refine the roots by a bisection algorithm; faster techniques are discussed below.

The second heuristic is that if  $h = f^2$  has small trace then the eigenvalues should be near the zeros of  $f$ . More precisely, for any  $f(t) = \mathbf{x} \cdot \mathbf{y}(t)$  we have  $|f(r_n)| \leq \sqrt{Q(\mathbf{x})}$  for every  $n$ ; this gives an effective bound on the possible locations of  $r_n$ , around zeros  $r_0$  of  $f$ . The strength of that bound is controlled by the size of  $|f'(r_0)|/\sqrt{Q(\mathbf{x})}$ , i.e. if this quantity is large then there is only a small interval around  $r_0$  that could possibly contain an eigenvalue. Since  $|f'(r_0)|/\sqrt{Q(\mathbf{x})}$  is unchanged if  $f$  is replaced by  $cf$  for a scalar  $c \neq 0$ , we need to fix a normalization; a convenient choice is to fix the derivative  $f'(r) = 1$  for an arbitrary  $r$  and optimize  $Q$ . Typically, the resulting optimal function will not have a zero at  $r$ ; our estimates for the eigenvalues are the special points where that happens. Precisely, let  $\tilde{V}_r$  denote the set of functions  $h = f^2$  of the form (4.2), constrained by  $f'(r) = 1$ , and let  $\tilde{h}_r = \tilde{f}_r^2 \in \tilde{V}_r$  have minimal trace. Then we look for eigenvalues at the points  $r$  such that  $\tilde{f}_r(r) = 0$ .

Fortunately, this heuristic is equivalent to the one described above, in the sense that they produce the same estimates, as the following lemma shows.

**Lemma 4.2.**

$$(4.6) \quad q'(r) = \frac{\tilde{h}'_r(r)}{\sum_{n=1}^{\infty} \tilde{h}_r(r_n)}.$$

*Proof.* Let  $\tilde{\mathbf{x}}(r)$  be the coefficient vector corresponding to  $\tilde{h}_r$ , and set  $\tilde{q}(r) = Q(\tilde{\mathbf{x}}(r))^{-1}$ . Then, as in (4.4), we have  $\tilde{q}(r) = \mathbf{y}'(r)^T A^{-1} \mathbf{y}'(r)$  and  $\tilde{\mathbf{x}}(r) = A^{-1} \frac{\mathbf{y}'(r)}{\tilde{q}(r)}$ . Next, we rewrite (4.5) as  $q'(r) = 2\mathbf{y}(r)^T A^{-1} \mathbf{y}'(r) = 2\tilde{q}(r)\mathbf{y}(r) \cdot \tilde{\mathbf{x}}(r) = 2\tilde{q}(r)\tilde{f}_r(r) = \tilde{q}(r)h'_r(r)$ .  $\square$

*Remarks 4.3.*

- (1) Note that  $\tilde{h}'_r(r) = 2\tilde{f}_r(r)$ . Thus, the zeros of  $q'(r)$  coincide with those of  $\tilde{f}_r(r)$ .
- (2) Equation (4.3) could be written  $q'(r) = \frac{h'_r(r)}{\sum_n h_r(r_n)}$ . The similarity with (4.6) is due to the symmetry of the inner product associated to  $A^{-1}$ , as the proof above shows.

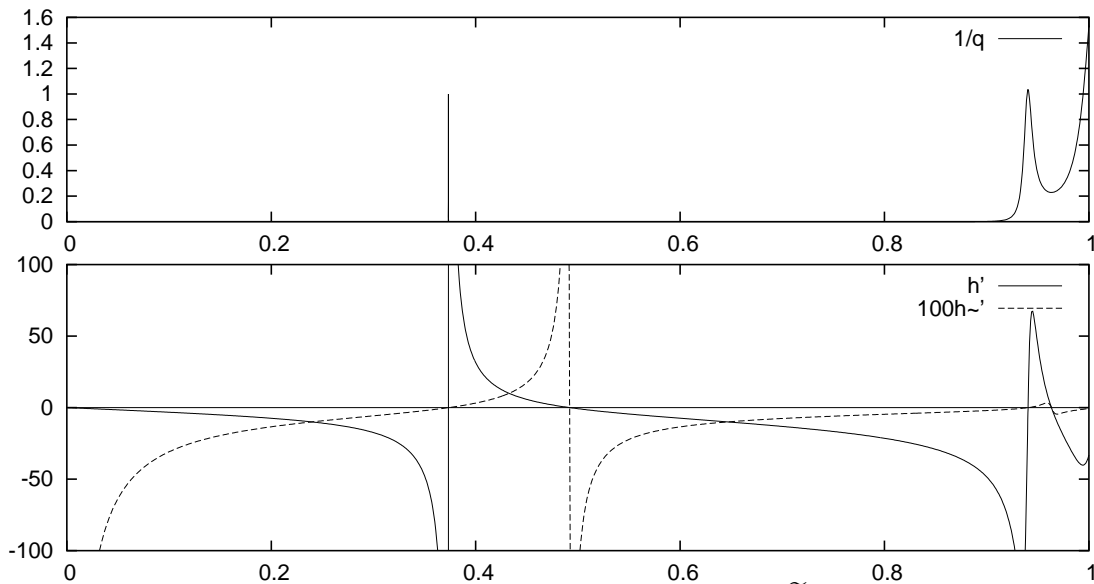


FIGURE 4.1. Plots of  $1/q(r)$  (top),  $h'_r(r)$  and  $100\tilde{h}'_r(r)$  (bottom)

Figure 4.1 shows the graph of  $q(r)^{-1} \gtrsim F_X(r)$  for the even forms on  $\Gamma_0(127)$  with trivial character, with  $X = 36$  and  $M = 3000$  sample points. Recall that the eigenvalues must lie in the intervals where  $F_X(r) \geq 1$ . The fact that we see a spike of height just above 1 indicates not only that the method works, but also that our trace formula and class number computations are likely correct. Similar conclusions apply to every pair  $N, \chi$  that we tested.

Below the plot of  $q(r)^{-1}$  we have superimposed the graphs of  $h'_r(r)$  and  $100\tilde{h}'_r(r)$ . While both functions are everywhere differentiable,  $\tilde{h}'_r(r)$  behaves much more stably around the maxima of  $q(r)^{-1}$ , as the graph shows. (Conversely,  $h'_r(r)$  is more stable around the minima.) Thus, while either function may be used for a bisection algorithm,  $\tilde{h}_r$  is preferable if one wishes to apply Newton's method. A simple iterative technique that seems to converge almost as quickly as Newton's method, yet uses only first derivative information, is to replace  $r$  by the nearest zero of  $\tilde{f}_r$  (which may be located by bisection without any additional matrix operations).

Using this technique, we get the estimate  $r = 0.3733851150$  for the first eigenvalue, and prove moreover that it lies in the interval  $(0.3733851149, 0.3733851152)$ . In fact, one can show that there must be an eigenvalue in that interval, by computing the trace formula for a test function which is positive up to 0.9 (just below the spike from the second eigenvalue), and non-positive everywhere else; such a function is easily constructed with Fourier transform of small compact support. We remark also that our estimate agrees to within  $10^{-8}$  with the value  $r = 0.37338511$  computed by Strömberg [36, 35] using a version of Hejhal's heuristic algorithms adapted to congruence subgroups. In a forthcoming paper [10], we combine the techniques of this section with those of [35] to study the distribution of eigenvalues in the level aspect.

**4.2. Application to  $\Gamma(1)$ .** Applying the method described above to  $\Gamma(1)$ , we obtain the estimates for the first ten  $r$ -values shown in Table 4.1, rounded to nine significant figures. These agree with the results of [11], where the eigenvalues are computed and proven correct to 100 places, using a high-precision version of Hejhal's algorithms and a certification technique. What was not shown in [11] is that the list is complete. Here we establish that fact with the trace formula.

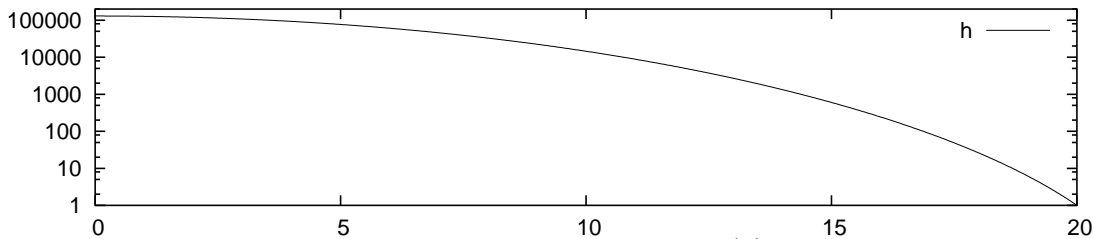
$n$	$r_n$	$n$	$r_n$
1	9.53369526	6	16.6442592
2	12.1730083	7	17.7385634
3	13.7797514	8	18.1809178
4	14.3585095	9	19.4234815
5	16.1380732	10	19.4847139

TABLE 4.1. First ten  $r$ -values on  $\Gamma(1)$

**Proposition 4.4.** *The eigenvalues listed in Table 4.1 are all simple. There are no other eigenvalues with  $r \leq 20$  on  $\Gamma(1)$ .*

*Proof.* Since the eigenvalues listed in Table 4.1 are known to exist and are given very precisely in [11], we may remove their contribution from the trace formula. It then suffices to exhibit a test function  $h(r)$  which is non-negative for  $r \in \mathbb{R}$  (since Selberg's conjecture is true for  $\Gamma(1)$ ) and  $\geq 1$  for  $|r| \leq 20$ , yet such that the trace (over the remaining eigenvalues) is  $< 1$ . We



FIGURE 4.2. Plot of  $h_{20}$  for  $\Gamma(1)$ 

choose the optimal function  $h_{20}$  with  $X = 1$  and  $M = 200$  sample points, which is graphed on a logarithmic scale in Figure 4.2. The trace over higher eigenvalues is less than 0.9. We remark that the support  $[-1, 1]$  is small enough that there is no parabolic or hyperbolic contribution to the trace formula.  $\square$

**4.3. Application to Selberg's conjecture.** By the discussion above, in order to demonstrate Selberg's conjecture for a particular  $\langle \Gamma_0(N), \chi \rangle$ , it suffices to show that  $F_X(r) < 1$  for  $r \in i[-\frac{1}{2}, \frac{1}{2}]$  (or even just  $r \in i[-\frac{7}{64}, \frac{7}{64}]$ , in light of the known lower bound). In principle this could be done by carrying out the above procedure using interval arithmetic for a large choice of  $X$ . In practice, it is enough to find the optimal test function  $h_0$  for  $r = 0$ . In all cases known to us, it was then possible to exhibit an  $M' < M$  such that the corresponding coefficients  $x_n$  satisfy  $x_n < 0$  for  $n \leq M'$  and  $x_n > 0$  for  $n > M'$ ; thus, for  $\nu \in \mathbb{R}$ ,

$$h_0(i\nu) = \left( \frac{\sinh \delta\nu/2}{\delta\nu/2} \right)^4 \left( \sum_{n=0}^{M-1} x_n \cosh \delta n\nu \right)^2 \geq \left( \sum_{n=0}^{M-1} x_n \cosh(\delta M'\nu) \right)^2 = \cosh^2(\delta M'\nu) \geq 1.$$

Therefore, if the trace is  $< 1$ , there cannot be any exceptional eigenvalues. In this way, a single test function suffices to show Selberg's conjecture.

To prove Theorem 1, we wrote two programs [8, trace.c and opt.c]. The first computed the trace formula for a given  $N$  and *all*  $\chi$  for a fixed test function  $h$  of the form  $\left( \frac{\sin At \sin Bt}{At Bt} \right)^2$ ,  $A + B \leq 18$ . This program worked for the majority of cases. The remaining ones were handled by the second program, which carried out the minimization procedure for a single  $(N, \chi)$ . There were twelve exceptional cases due to Galois representations, at the levels given in Table 5.1; these will be treated in the next section. At level 857 we were unable to show the conjecture for one character which cannot possibly admit a Galois representation. Most likely the conjecture is true for  $\Gamma_1(857)$  as well, but our list of class numbers is not sufficient to show it.

## 5. GALOIS REPRESENTATIONS

**5.1. General facts for non-dihedral representations.** It turns out that the characterization of 2-dimensional Galois representations of prime conductor given by Serre [34] and Vignéras [44] can be carried over in a straightforward way to yield a convenient and precise characterization of those Galois extensions of  $\mathbb{Q}$  which give rise to non-dihedral (even) representations of squarefree conductor.

Assume that  $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}(2, \mathbb{C})$  ( $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ ) is an irreducible, even, non-dihedral Galois representation of squarefree conductor  $N$ . Let  $\varepsilon = \det \rho$  and let  $\tilde{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{PGL}(2, \mathbb{C})$  be the composition of  $\rho$  with the projection  $\mathrm{GL}(2, \mathbb{C}) \rightarrow \mathrm{PGL}(2, \mathbb{C})$ . We may identify  $\tilde{\rho}(G_{\mathbb{Q}})$  with  $A_4$ ,  $S_4$  or  $A_5$ . Now for each prime  $p \mid N$ , if  $I_p < G_{\mathbb{Q}_p} < G_{\mathbb{Q}}$  is the inertia and decomposition subgroup for a place of  $\overline{\mathbb{Q}}$  over  $p$ , then since  $\mathrm{ord}_p N = 1$  we have  $\rho|_{I_p} = \psi \oplus 1$

for some 1-dimensional representation  $\psi \neq 1$  of  $I_p$ . Hence the canonical homomorphisms  $\rho(I_p) \rightarrow \varepsilon(I_p)$  and  $\rho(I_p) \rightarrow \tilde{\rho}(I_p)$  are in fact isomorphisms. But  $\varepsilon(I_p)$  is a cyclic subgroup of  $\mathbb{C}^\times$ , and  $\tilde{\rho}(I_p)$  is a subgroup of  $A_4$ ,  $S_4$  or  $A_5$ ; hence  $\varepsilon(I_p) \cong \rho(I_p) \cong \tilde{\rho}(I_p)$  is cyclic of order 2, 3, 4 or 5. The Artin map  $\mathbb{Q}_A^\times \rightarrow G_{\mathbb{Q}}^{\text{ab}}$  restricts to the local Artin map  $\mathbb{Q}_p^\times \rightarrow G_{\mathbb{Q}_p}^{\text{ab}}$ . Hence when viewing  $\varepsilon$  as a Dirichlet character on  $\mathbb{Z}/N\mathbb{Z}$  via class field theory,  $\varepsilon_p$  on  $(\mathbb{Z}/p\mathbb{Z})^\times$  is determined from  $\varepsilon|_{I_p}$ , and in particular  $\varepsilon_p$  and  $\varepsilon|_{I_p}$  have the same order. It follows that the conductor of  $\varepsilon$  is exactly equal to  $N$ , and in particular  $N$  is odd.

Let  $N_j$  be the product of primes  $p \mid N$  such that  $|\varepsilon(I_p)| = j$ , for  $j = 2, 3, 4, 5$ , so that  $N = N_2 N_3 N_4 N_5$ . Then for each  $p \mid N_j$  the character  $\varepsilon_p$  on  $(\mathbb{Z}/p\mathbb{Z})^\times$  is of order  $j$ ; thus  $p \equiv 1 \pmod{j}$ , and if  $j = 2$  then  $\varepsilon_p = (\cdot/p)$ , the Legendre symbol. We fix a subgroup  $H$  in  $\tilde{\rho}(G_{\mathbb{Q}})$  of index 4 if  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_4$  or  $S_4$ ; in the last case  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_5$  we instead let  $H$  be a subgroup of index 5. Let  $E_0/\mathbb{Q}$  be the quartic or quintic extension corresponding to  $\tilde{\rho}^{-1}(H)$ . Note that in each case the choice of  $H$  is uniquely determined up to conjugation, and the Galois closure of  $E_0$ , which we call  $E$ , is the fixed field of the kernel of  $\tilde{\rho}$ ; thus  $\tilde{\rho}(G_{\mathbb{Q}}) \cong \text{Gal}(E/\mathbb{Q})$ . Note that since  $\varepsilon$  is even,  $E$  is totally real, and thus  $E_0$  is *totally real*. Note also that by definition of the Artin conductor, each prime  $p \nmid N$  is unramified in  $E$  and hence in  $E_0$ . We now have:

1) If  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  then  $N = N_2 N_3$ , each prime  $p \mid N_2$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2^2$  and each prime  $p \mid N_3$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^3 \mathfrak{q}_2$ . (Here and below,  $\mathfrak{q}_1, \mathfrak{q}_2, \dots$  are always assumed to be pairwise distinct prime ideals in  $\mathcal{O}_{E_0}$ .)

2) If  $\text{Gal}(E/\mathbb{Q}) \cong S_4$  then  $N = N_2 N_3 N_4$ , each prime  $p \mid N_2$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2 \mathfrak{q}_3$ ,  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2^2$  or  $(p) = \mathfrak{q}_1^2$ ; each prime  $p \mid N_3$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^3 \mathfrak{q}_2$ , and each prime  $p \mid N_4$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^4$ .

3) If  $\text{Gal}(E/\mathbb{Q}) \cong A_5$  then  $N = N_2 N_3 N_5$ , each prime  $p \mid N_2$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2^2 \mathfrak{q}_3$ ; each prime  $p \mid N_3$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^3 \mathfrak{q}_2 \mathfrak{q}_3$ , and each prime  $p \mid N_5$  splits in  $E_0$  as  $(p) = \mathfrak{q}_1^5$ .

To prove this, we identify  $\tilde{\rho}(G_{\mathbb{Q}}) \cong \text{Gal}(E/\mathbb{Q})$ ; then  $D = \tilde{\rho}(G_{\mathbb{Q}_p})$  and  $I = \tilde{\rho}(I_p)$  are the decomposition and inertia subgroups for some prime ideal  $\mathfrak{p}$  in  $\mathcal{O}_E$  above  $(p)$ . Since  $\text{ord}_p N = 1$ ,  $D$  is cyclic (cf. [34, §6.3]). Now the ramification behavior of  $(p)$  in  $E_0$  is completely determined by the subgroups  $H, D, I$  of  $\text{Gal}(E/\mathbb{Q})$ , in fact already from knowing the conjugacy class in  $\text{Gal}(E/\mathbb{Q})$  of each of  $H, D, I$ . [To wit,  $\sigma \mapsto \mathcal{O}_{E_0} \cap \sigma \mathfrak{p}$  gives a surjection from  $\text{Gal}(E/\mathbb{Q})$  onto the set of prime ideals in  $\mathcal{O}_{E_0}$  above  $(p)$ ; here  $\mathcal{O}_{E_0} \cap \sigma \mathfrak{p} = \mathcal{O}_{E_0} \cap \sigma' \mathfrak{p}$  if and only if the double cosets  $H\sigma D$  and  $H\sigma' D$  are equal, and the ramification degree of  $\mathcal{O}_{E_0} \cap \sigma \mathfrak{p}$  above  $(p)$  is  $|I : I \cap \sigma^{-1} H \sigma|$ .] The proof can now be concluded by a straightforward case by case analysis, cf. [8, ramification.g] for a GAP program which carries this out. This concludes the proof.

We now prove a converse statement:

**Proposition 5.1.** *Let  $E_0$  be a quartic or quintic totally real number field such that its Galois closure  $E/\mathbb{Q}$  has Galois group  $\text{Gal}(E/\mathbb{Q}) \cong A_4, S_4$  or  $A_5$ . Let  $\tilde{\rho}$  be the composition of the canonical surjection of  $G_{\mathbb{Q}}$  onto  $\text{Gal}(E/\mathbb{Q})$  and an imbedding of  $\text{Gal}(E/\mathbb{Q})$  in  $\text{PGL}(2, \mathbb{C})$ . Assume that the product of those primes at which  $E/\mathbb{Q}$  is ramified can be expressed as  $N = N_2 N_3 N_4 N_5$  such that one of the ramification descriptions 1), 2) or 3) above holds. If  $\text{Gal}(E/\mathbb{Q}) \cong S_4$  then for each  $p \mid N_2$  with  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2^2$  or  $\mathfrak{q}_1^2$  in  $E_0$  we assume that the decomposition subgroup for  $(p)$  in  $E$  is cyclic. We also assume that for  $j = 2, 3, 4, 5$ , each prime  $p \mid N_j$  satisfies  $p \equiv 1 \pmod{j}$ . Then  $\tilde{\rho}$  admits a lifting  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}(2, \mathbb{C})$  with Artin conductor  $N$  and with  $\det \rho$  even.*

*Proof.* Note that  $E_0$  must be quartic if  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  or  $S_4$ , and quintic if  $\text{Gal}(E/\mathbb{Q}) \cong A_5$ , since  $A_4, S_4$  do not have any subgroups of index 5 and  $A_5$  does not have any subgroup of index 4. Also note that  $N$  is odd, since  $2 \not\equiv 1 \pmod{j}$  for  $j = 2, 3, 4, 5$ .

We first claim that each prime  $p \mid N$  is *tamely* ramified in  $E$ , viz.  $p \nmid \#\tilde{\rho}(I_p)$ . This is trivial unless  $p = 3$  or  $5$ , since  $\tilde{\rho}(I_p)$  is isomorphic to a subgroup of  $A_4, S_4$  or  $A_5$ . Assume  $3 \mid \#\tilde{\rho}(I_3)$ . Then, examining all possibilities (up to conjugacy) for subgroups  $I = \tilde{\rho}(I_3) \subset D = \tilde{\rho}(G_{\mathbb{Q}_p})$  of  $\tilde{\rho}(G_{\mathbb{Q}})$  with  $3 \mid \#I$  and  $I$  normal in  $D$ , we find that (3) cannot ramify in  $E_0$  as described by 1), 2) or 3), a contradiction. (The details of this case by case study are carried out by the GAP program [8, ramification.g].) Hence  $3 \nmid \#\tilde{\rho}(I_3)$ . Similarly  $5 \nmid \#\tilde{\rho}(I_5)$ .

Now fix any prime  $p \mid N$ . Since  $p$  is tamely ramified in  $E$  the inertia group  $I = \tilde{\rho}(I_p)$  is a cyclic subgroup of  $\tilde{\rho}(G_{\mathbb{Q}}) \cong \text{Gal}(E/\mathbb{Q})$ . Since  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_4, S_4$  or  $A_5$  this implies that  $\#I = 2, 3, 4$  or  $5$ . Also the decomposition group  $D = \tilde{\rho}(G_{\mathbb{Q}_p})$  contains  $I$  and is contained in the normalizer of  $I$ . If  $\#I \geq 3$  this implies that  $D = I$  or  $D$  is dihedral of order  $2 \cdot \#I$ , and examining case by case (cf. [8, ramification.g]) we find that  $(p)$  must ramify in  $E_0$  as follows:

$$\begin{aligned} \text{If } \#I = 3: \quad & [\tilde{\rho}(G_{\mathbb{Q}}) \cong A_4 \text{ or } S_4] \implies (p) = \mathfrak{q}_1^3 \mathfrak{q}_2; \\ & \tilde{\rho}(G_{\mathbb{Q}}) \cong A_5 \implies [(p) = \mathfrak{q}_1^3 \mathfrak{q}_2 \mathfrak{q}_3 \text{ or } (p) = \mathfrak{q}_1^3 \mathfrak{q}_2^2]. \\ \text{If } \#I = 4 \text{ (and thus } \tilde{\rho}(G_{\mathbb{Q}}) \cong S_4): & \quad (p) = \mathfrak{q}_1^4. \\ \text{If } \#I = 5 \text{ (and thus } \tilde{\rho}(G_{\mathbb{Q}}) \cong A_5): & \quad (p) = \mathfrak{q}_1^5. \end{aligned}$$

Comparing this with our assumptions, we find that  $p \mid N_j$  must hold for  $j = \#I$  (and also we may exclude the possibility  $(p) = \mathfrak{q}_1^3 \mathfrak{q}_2^2$  above). Hence our assumptions also imply  $p \equiv 1 \pmod{\#I}$ , and this forces  $D = I$  and shows that  $\tilde{\rho}|_{G_{\mathbb{Q}_p}}$  has a lifting  $\rho : G_{\mathbb{Q}_p} \rightarrow \text{GL}(2, \mathbb{C})$  of conductor  $p$  (cf. [34, p. 248 (Lemma)]).

Now assume  $\#I = 2$ . Then if  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_4$  or  $A_5$ ,  $I$  must be conjugate to  $\langle (12)(34) \rangle$ , and thus  $D = I$  or  $D = \langle (12)(34), (13)(24) \rangle$ . However the second possibility implies  $(p) = \mathfrak{q}_1^2$  (if  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_4$ ) or  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_3$  (if  $\tilde{\rho}(G_{\mathbb{Q}}) \cong A_5$ ), which is impossible by our assumptions. Hence  $D = I$ . Finally assume  $\tilde{\rho}(G_{\mathbb{Q}}) \cong S_4$ ; then there are two possibilities for  $I$  up to conjugacy. The first case is  $I = \langle (12) \rangle$ ; then either  $D = I$ , which implies  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2 \mathfrak{q}_3$ , or else  $D = \langle (12), (34) \rangle$ , which implies  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2$ . Hence by our assumptions we must have  $D = I$ , and  $\tilde{\rho}|_{G_{\mathbb{Q}_p}}$  has a lifting of conductor  $p$ . The second case is  $I = \langle (12), (34) \rangle$ , giving five possibilities for  $D$  and giving  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2^2$  or  $(p) = \mathfrak{q}_1^2$ . In this case we refer to the explicit assumption made in our formulation of the proposition to see that  $D$  is cyclic. Hence in all cases,  $D$  is cyclic, and thus  $\tilde{\rho}|_{G_{\mathbb{Q}_p}}$  again has a lifting of conductor  $p$  (cf. [34, §6.3]).

Hence by Tate's theorem [34, Theorem 5] (and [34, p. 227]),  $\tilde{\rho}$  indeed has a lifting  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}(2, \mathbb{C})$  with Artin conductor  $N$ . Since  $E$  is totally real,  $\det \rho$  is even.  $\square$

Note that if  $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}(2, \mathbb{C})$  is *one* lifting of  $\tilde{\rho}$  with Artin conductor  $N$ , and  $\varepsilon = \det \rho$ , then by [34, Theorem 5] and [23, p. 10 (bottom)], *all* liftings with conductor  $N$  are obtained as  $\rho \otimes \varepsilon_A^{-1}$ , where  $A$  runs through the divisors of  $N$ . Since  $\rho$  is non-dihedral, the  $d(N)$  Galois representations  $\rho \otimes \varepsilon_A^{-1}$  for  $A \mid N$  are easily shown to be pairwise non-isomorphic. Note here  $\det(\rho \otimes \varepsilon_A^{-1}) = \varepsilon \cdot \varepsilon_A^{-2} = \varepsilon_{N/A} \cdot \varepsilon_A^{-1} = \varepsilon^{[A]}$  (cf. (3.10)). Hence since  $\varepsilon_p(\mathbb{Z}/p\mathbb{Z})$  is of order  $j$  for each  $p \mid N_j$  we conclude that if  $\rho$  runs through all liftings as above then  $\det \rho$  runs through all Dirichlet characters  $\chi \pmod{N}$  such that  $\chi_p$  is of order  $j$  for each  $p \mid N_j$ ,  $j = 2, 3, 4$ , and  $\chi_p \in \{\varepsilon_p, \overline{\varepsilon_p}\}$ , for each  $p \mid N_5$  (the last condition is only relevant in the icosahedral case). We also see that each such Dirichlet character is obtained exactly  $d(N_2)$  times. But there is only one imbedding each of  $A_4$  and  $S_4$  in  $\text{PGL}(2, \mathbb{C})$  up to conjugacy;

hence if  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  or  $S_4$  then the above  $d(N)$  liftings of  $\tilde{\rho}$  form the complete list of non-isomorphic Galois representations  $\tilde{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}(2, \mathbb{C})$  such that  $\ker(\tilde{\rho})$  has fixed field  $E$ .

Now assume  $\text{Gal}(E/\mathbb{Q}) \cong A_5$ . Then there are exactly two conjugacy classes of imbeddings of  $A_5$  in  $\text{PGL}(2, \mathbb{C})$ ; choose  $\tilde{\rho}' : G_{\mathbb{Q}} \rightarrow \text{PGL}(2, \mathbb{C})$  so that  $\tilde{\rho}$  and  $\tilde{\rho}'$  correspond to non-conjugate imbeddings of  $\text{Gal}(E/\mathbb{Q})$  in  $\text{PGL}(2, \mathbb{C})$ . For each  $\sigma \in G_{\mathbb{Q}}$  such that  $\tilde{\rho}(\sigma)$  is of order 5, we define  $J(\sigma) \in \{1, 2\}$  such that the ratio of the two eigenvalues of  $\tilde{\rho}(\sigma)$  (in appropriate order) equals  $e(J(\sigma)/5)$ . Then a simple explicit study of the two  $A_5$ -imbeddings shows that the ratio of the two eigenvalues of  $\tilde{\rho}'(\sigma)$  is  $e((3 - J(\sigma))/5)$ , for any such  $\sigma \in G_{\mathbb{Q}}$ . Fix any  $p \mid N_5$  and choose  $x \in \mathbb{Z}_p^\times$  which projects to a generator of  $(\mathbb{Z}/p\mathbb{Z})^\times$ . Let  $\omega_p : \mathbb{Q}_p^\times \rightarrow G_{\mathbb{Q}_p}^{\text{ab}}$  be the local Artin map. Since  $x \in \mathbb{Z}_p^\times$  there is  $\sigma \in I_p$  such that  $\omega_p(x) = \sigma G'_{\mathbb{Q}_p}$  in  $G_{\mathbb{Q}_p}^{\text{ab}} = G_{\mathbb{Q}_p}/G'_{\mathbb{Q}_p}$ . By class field theory we view  $\varepsilon$  both as a character on  $G_{\mathbb{Q}}$  and a character on  $\mathbb{Q}_A^\times/\mathbb{Q}^\times$  trivial on  $\mathbb{R}^\times$ . Then  $\varepsilon(\sigma) = \varepsilon(x) \neq 1$  and since the determinant map  $\rho(I_p) \rightarrow \varepsilon(I_p)$  is an isomorphism,  $\rho(\sigma)$  is a matrix of order 5 generating  $\rho(I_p)$ . Recall  $\rho|_{I_p} \cong \psi \oplus 1$  for some 1-dimensional representation  $\psi$  of  $I_p$ ; hence we may choose basis in  $\mathbb{C}^2$  so that  $\rho(\sigma) = \begin{pmatrix} e(k/5) & 0 \\ 0 & 1 \end{pmatrix}$  for some  $k \in \{1, 2, 3, 4\}$ , and we see  $J(\sigma) \equiv \pm k \pmod{5}$  and  $\varepsilon(\sigma) = \det \rho(\sigma) = e(k/5)$ . By the same argument, if  $\rho' : G \rightarrow \text{GL}(2, \mathbb{C})$  is a lifting of  $\tilde{\rho}'$  and  $\varepsilon' = \det \rho'$ , then using that the ratio of the two eigenvalues of  $\tilde{\rho}'(\sigma)$  is  $e((3 - J(\sigma))/5)$ , we find  $\varepsilon'(x) = e(k'/5)$  for some  $k' \in \{1, 2, 3, 4\}$  with  $k' \equiv \pm 2k \pmod{5}$ . Thus  $\varepsilon'_p = \varepsilon_p^2$  or  $\varepsilon_p^3$ , and this holds for each  $p \mid N_5$ . Hence as above we see that if  $\rho'$  runs through all the  $d(N)$  liftings of  $\tilde{\rho}'$  with conductor  $N$  then  $\det \rho'$  runs through all Dirichlet characters  $\chi \pmod{N}$  such that  $\chi_p$  is of order  $j$  for each  $p \mid N_j$ ,  $j = 2, 3, 4$ , and  $\chi_p \in \{\varepsilon_p^2, \varepsilon_p^3\}$ , for each  $p \mid N_5$ , and each such Dirichlet character is obtained exactly  $d(N_2)$  times. Clearly, no lifting of  $\tilde{\rho}'$  can be isomorphic to a lifting of  $\tilde{\rho}$ . Hence, if  $\text{Gal}(E/\mathbb{Q}) \cong A_5$  then there are exactly  $2d(N)$  non-isomorphic Galois representations  $\tilde{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}(2, \mathbb{C})$  such that  $\ker(\tilde{\rho})$  has fixed field  $E$ ; a complete list of representatives for these is formed by the above  $d(N)$  liftings of  $\tilde{\rho}$  together with the  $d(N)$  liftings of  $\tilde{\rho}'$ .

*Remark 5.2.* We point out some more restrictions which are valid whenever the situation in Proposition 5.1 holds. For each  $p \mid N_3 N_5$  the character  $\varepsilon_p$  is of odd order and thus  $\varepsilon_p(-1) = 1$ ; hence  $(-1)^{\frac{N_2-1}{2}} (-1)^{\frac{N_4-1}{4}} = \varepsilon_{N_2}(-1) \varepsilon_{N_4}(-1) = \varepsilon(-1) = 1$  (and hence if  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  or  $A_5$  we must have  $N_2 \equiv 1 \pmod{4}$ ). Some restrictions can also be derived from the fact that there cannot exist a normal subgroup  $H_0 \subsetneq \tilde{\rho}(G_{\mathbb{Q}})$  which contains  $\tilde{\rho}(I_p)$  for all  $p \mid N$ . In particular, if  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  then  $N_3 > 1$ , and if  $\text{Gal}(E/\mathbb{Q}) \cong S_4$  then  $N_2 N_4 > 1$ .

*Remark 5.3.* Note that the discriminant  $d$  of  $E_0$  can be computed from the ramification descriptions (of course  $d > 0$  since  $E_0$  is totally real): If  $\text{Gal}(E/\mathbb{Q}) \cong A_4$  then  $d = N^2$ . If  $\text{Gal}(E/\mathbb{Q}) \cong S_4$  then  $d = N'_2 N_2''^2 N_3^2 N_4^3$  where  $N'_2$  is the product of primes  $p \mid N_2$  which factor as  $(p) = \mathfrak{q}_1^2 \mathfrak{q}_2 \mathfrak{q}_3$  in  $E_0$ , and  $N_2'' = N_2/N'_2$ . If  $\text{Gal}(E/\mathbb{Q}) \cong A_5$  then  $d = N_2^2 N_3^2 N_5^4$ .

**5.2. Explicit examples.** Recall that our computations with the trace formula, after removing the contribution from dihedral forms, allowed us to prove that the Selberg eigenvalue conjecture holds true for any  $(N, \chi)$  with  $N$  squarefree  $< 857$ , except possibly for those  $(N, \chi)$  listed in Table 5.1 below.<sup>4</sup> On the other hand, for each listed  $(N, \chi)$  it is possible to

<sup>4</sup>In Tables 5.1–5.3 we list the prime factorization of each  $N$ , say  $N = \prod_k p_k$  with  $p_1 < p_2 < \dots < p_r$ , and we specify the type of  $\chi$  by a vector  $o_1, \dots, o_r$  of numbers in  $\{2, 3, 4, 5\}$ , meaning that  $\chi_{p_k}$  on  $(\mathbb{Z}/p_k\mathbb{Z})^\times$  is of order  $o_k$  for each  $k$ .

$N$	$\chi$	$G$	Polynomial defining $E_0$	$N$	$\chi$	$G$	Polynomial defining $E_0$
163	3	$A_4$	$x^4 - x^3 - 7x^2 + 2x + 9$	679 = 7 · 97	3,3	$A_4$	$x^4 - x^3 - 24x^2 + 19x + 117$
277	3	$A_4$	$x^4 - x^3 - 11x^2 + 4x + 12$	703 = 19 · 37	3,3	$A_4$	$x^4 - 2x^3 - 19x^2 + 19x + 19$
349	3	$A_4$	$x^4 - x^3 - 10x^2 + 3x + 20$	709	3	$A_4$	$x^4 - 17x^2 - 13x + 35$
397	3	$A_4$	$x^4 - 13x^2 - 2x + 19$	763 = 7 · 109	3,3	$A_4$	$x^4 - x^3 - 16x^2 + 17x + 38$
547	3	$A_4$	$x^4 - 21x^2 - 3x + 100$	785 = 5 · 157	4,4	$S_4$	$x^4 - x^3 - 294x^2 - 1521x + 3186$
607	3	$A_4$	$x^4 - 2x^3 - 13x^2 + 7x + 33$	853	3	$A_4$	$x^4 - x^3 - 28x^2 + 31x - 2$

TABLE 5.1. Non-dihedral Galois representations of squarefree conductor  $N < 857$ .

$N$	$\chi$	$N$	$\chi$
2067 = 3 · 13 · 53	2, 2, 4	2611 = 7 · 373	3, 3
2143	3	2753	4
2221	5 ( $\chi(2) = e(\pm\frac{1}{5})$ )	2767	3
2341	5 ( $\chi(7) = e(\pm\frac{3}{5})$ )	2791	5 ( $\chi(6) = e(\pm\frac{2}{5})$ )
2381	5 ( $\chi(3) = e(\pm\frac{1}{5})$ )	2827 = 11 · 257	5, 4 ( $\chi_{11}(2) = e(\pm\frac{1}{5})$ )
2545 = 5 · 127	4, 4	2863 = 7 · 409	3, 4
2591	5 ( $\chi(7) = e(\pm\frac{1}{5})$ )	2993 = 41 · 73	5, 4 ( $\chi_{41}(6) = e(\pm\frac{1}{5})$ )

TABLE 5.2. Exceptional cases

construct a tetrahedral or octahedral Galois representation with conductor  $N$  and determinant character equal to  $\chi$ ; we list in Table 5.1 the polynomials defining quartic number fields  $E_0$  from which the Galois representation can be deduced as in Proposition 5.1.

In view of the results of Langlands [25] and Tunnell [41] (and the elementary discussion of multiplicity in Section 5.1), for each such  $(N, \chi)$  there exists a Maass waveform with eigenvalue exactly  $\lambda = \frac{1}{4}$  on  $\langle \Gamma_0(N), \chi \rangle$  (orthogonal to any possible CM-form). With this information added, our trace formula computations now allow us to conclude the main result of this paper (a restatement of Theorem 1 in the introduction):

**Theorem.** *The Selberg conjecture is true on any  $\langle \Gamma_0(N), \chi \rangle$  with  $N$  squarefree  $< 857$ .*

*Remark 5.4.* If we only used the trace formula for the full spaces of Maass waveforms with respect to  $\langle \Gamma_0(N), \chi \rangle$ , without separating the even and odd forms as in Section 2, then we would only be able to prove the corresponding result for  $N < 647$ .

Restricting to characters  $\chi$  such that  $o(\chi_p) \in \{2, 3, 4, 5\}$  for all  $p \mid N$ , the trace formula computations can be extended to larger squarefree  $N$ : We were able to prove the Selberg eigenvalue conjecture for all such pairs  $(N, \chi)$  with  $857 \leq N < 3000$  except those listed in Table 5.3, and fourteen more exceptional cases, as shown in Table 5.2.

On the other hand, for each  $(N, \chi)$  in Table 5.3, there exists a Galois representation with conductor  $N$  and determinant character equal to  $\chi$ ; we list polynomials defining these (as in Table 5.1). Using Langlands [25] and Tunnell [41] (and a comparison of multiplicity in the trace formula results vs. Section 5.1), we are now able to conclude:

**Proposition 5.5.** *The Selberg conjecture is true on any  $\langle \Gamma_0(N), \chi \rangle$  with  $N$  squarefree  $< 3000$  and  $\chi$  satisfying  $o(\chi_p) \in \{2, 3, 4, 5\}, \forall p \mid N$ , except possibly for the two pairs  $(N, \chi)$  in Table 5.3 of type “ $A_5$ ”, and the fourteen exceptional cases in Table 5.2. (If Artin’s conjecture is true then the Selberg conjecture also holds for the two “ $A_5$ ” cases.)*

Most of the defining polynomials in Tables 5.1 and 5.3 (for the non- $A_5$  cases) can be found using number field tables [31]. The two  $A_5$  cases are known from the recent computations by Doud and Moore, [13]. The only other cases which cannot be found in existing tables

$N$	$\chi$	$G$	Polynomial defining $E_0$	$N$	$\chi$	$G$	Polynomial defining $E_0$
937	3	$A_4$	$x^4 - x^3 - 16x^2 + 11x + 54$	2135 = 5 · 7 · 61	2,3,3	$A_4$	$x^4 - 33x^2 - 20x + 16$
949 = 13 · 73	3,3	$A_4$	$x^4 - 19x^2 - 8x + 49$	2141	5	$A_5$	$x^5 - x^4 - 856x^3 + 4025x^2$ $+ 28501x - 40877$
995 = 5 · 199	2,3	$A_4$	$x^4 - x^3 - 22x^2 + 8x + 24$	2159 = 17 · 127	4,3	$S_4$	$x^4 - 51x^2 - 51x + 323$
1009	3	$A_4$	$x^4 - x^3 - 28x^2 + 59x + 4$	2279 = 43 · 53	3,2	$S_4$	$x^4 - x^3 - 11x^2 + 3x + 25$
1073 = 29 · 37	2,3	$A_4$	$x^4 - x^3 - 15x^2 + 4x + 45$	2311	2	$A_4$	$x^4 - 2x^3 - 29x^2 + 11x + 117$
1147 = 31 · 37	3,3	$A_4$	$x^4 - 2x^3 - 25x^2 - 3x + 45$	2353 = 13 · 181	3,3	$A_4$	$x^4 - 31x^2 - 5x + 116$
1267 = 7 · 181	3,2	$A_4$	$x^4 - x^3 - 17x^2 - 14x + 15$	2587 = 13 · 199	3,3	$A_4$	$x^4 - 2x^3 - 49x^2 + 55x - 12$
1267 = 7 · 181	3,3	$A_4$	$x^4 - x^3 - 31x^2 - 42x + 36$	2639 = 7 · 13 · 29	3,2,2	$A_4$	$x^4 - x^3 - 52x^2 - 21x + 64$
1333 = 31 · 43	3,3	$A_4$	$x^4 - x^3 - 22x^2 - x + 71$	2689	3	$A_4$	$x^4 - x^3 - 65x^2 + 256x - 192$
1343 = 17 · 79	2,3	$A_4$	$x^4 - 15x^2 - 17x + 1$	2713	4	$S_4$	$x^4 - x^3 - 1017x^2 + 1526x + 28052$
1345 = 5 · 269	4,4	$S_4$	$x^4 - x^3 - 504x^2 + 84x + 60856$	2743	3,3	$A_4$	$x^4 - x^3 - 31x^2 + 2x + 148$
1399	3	$A_4$	$x^4 - x^3 - 23x^2 + 18x + 68$	2777	4	$S_4$	$x^4 - x^3 - 1041x^2 + 7116x - 12247$
1699	3	$A_4$	$x^4 - x^3 - 28x^2 + 5x + 75$	2777	4	$S_4$	$x^4 - x^3 - 1041x^2 + 1562x + 93279$
1777	3	$A_4$	$x^4 - 2x^3 - 37x^2 + 10x + 29$	2777	2	$S_4$	$x^4 - x^3 - 4x^2 + x + 2$
1789	3	$A_4$	$x^4 - 31x^2 - 44x + 69$	2797	3	$A_4$	$x^4 - 35x^2 - 13x + 131$
1879	3	$A_4$	$x^4 - x^3 - 28x^2 + 13x + 88$	2803	3	$A_4$	$x^4 - x^3 - 52x^2 + 17x + 4$
1951	3	$A_4$	$x^4 - x^3 - 38x^2 + 25x + 36$	2857	4	$S_4$	$x^4 - x^3 - 1071x^2 - 1250x - 279$
1951	5	$A_5$	$x^5 - x^4 - 780x^3 - 1795x^2$ $+ 3106x + 344$	2881 = 43 · 67	3,3	$A_4$	$x^4 - x^3 - 41x^2 - 52x + 113$
1957 = 19 · 103	3,3	$A_4$	$x^4 - x^3 - 26x^2 - x + 107$	2945 = 5 · 19 · 31	2,3,3	$A_4$	$x^4 - x^3 - 56x^2 - 100x - 40$
2051 = 7 · 293	3,2	$A_4$	$x^4 - x^3 - 37x^2 - 18x + 31$	2977 = 13 · 229	3,3	$A_4$	$x^4 - x^3 - 48x^2 + 17x + 548$
2077 = 31 · 67	3,3	$A_4$	$x^4 - x^3 - 25x^2 + 24x + 115$	2983 = 19 · 157	3,2	$S_4$	$x^4 - x^3 - 8x^2 + 8x - 1$
2131	3	$A_4$	$x^4 - x^3 - 25x^2 + 10x + 123$	2993 = 41 · 73	4,4	$S_4$	$x^4 - x^3 - 1122x^2 - 8792x + 79864$

TABLE 5.3. Non-dihedral Galois representations of squarefree conductor  $857 \leq N < 3000$

(known to us) are the octahedral representations with  $N = 785, 1345, 2159, 2713, 2777$  ( $\chi$  of type 4), 2857 and 2993; for these we used a targeted Hunter search imitating [13] (using the precise factorization requirements noted in Section 5.1). We remark that the two polynomials specified for  $N = 2777$  and  $\chi$  of type 4 have different splitting fields, and thus lead to inequivalent Galois representations.

We also used an exhaustive Hunter search to prove that no non-dihedral Galois representations exist for the cases in Table 5.2 (whenever this could not be ruled out already by Proposition 5.1 or Remark 5.2). An alternative approach for the tetrahedral and octahedral cases requiring “no searching” would be via class field theoretic computations; in fact, this was kindly done for us at an early stage by D. Doud for the representation with  $N = 785$ .

We remark that the possibility of tetrahedral representations leading to some  $(N, \chi)$  as in Table 5.2 can alternatively be ruled out by using the number field tables in [31]; this is also true for icosahedrals at  $N = 2143$  and  $N = 2611$ . Also, for those cases in Table 5.2 where  $N$  is prime and  $\chi$  is of order 5, so that non-icosahedral representations are ruled out from the start, the non-existence of icosahedral representations follows from the computations of Doud and Moore [13].

We note the following consequence of our trace formula computations combined with the results of Langlands [25] and Tunnell [41], and the fact that we have shown that there are no Galois representations with data as in Table 5.2:

**Theorem 5.** *The complete list of even tetrahedral and octahedral Galois representations with squarefree Artin conductor  $N < 3000$  is given by Tables 5.1 and 5.3. If Artin’s conjecture is true, the only even icosahedral Galois representations with squarefree Artin conductor  $N < 3000$  occur for  $N = 1951$  and  $N = 2141$ , cf. Table 5.3.*

We remark that the computations of Doud and Moore [13] prove unconditionally that there are exactly six  $A_5$ -extensions of  $\mathbb{Q}$  which lead to even icosahedral Galois representations with Artin conductor  $N$  a *prime number*  $< 10000$ ; these occur for  $N = 1951, 2141, 3701, 3821, 8501, 9461$ . In a paper in preparation, [9], we will continue the study of finding all Galois representations of small conductor.

## 6. COMPLEXITY

There are two factors governing the complexity of demonstrating Selberg's conjecture for  $\langle \Gamma_0(N), \chi \rangle$  using the trace formula. First, as  $N$  increases, the density of high eigenvalues, given by Weyl's law, increases roughly in proportion to  $N$ ; thus, we need test functions  $h(t)$  that decay more rapidly at infinity for larger  $N$ . Second, the lowest eigenvalues encroach more and more closely upon the lower bound  $\frac{1}{4}$ ; heuristics based on the trace formula and numerical experiments [10] indicate that the lowest  $r$ -value for  $\langle \Gamma_0(N), \chi \rangle$  (in cases where Selberg's conjecture is satisfied) is typically of size  $N^{-1/3}$ . (See [10] for a more thorough discussion of this point.) Thus, our test functions  $h(t)$  should vary for  $t$  on the order of  $N^{-1/3}$  and decay sharply thereafter.

On the other hand, as was discussed briefly in Section 2.4, computing the trace formula with a test function whose Fourier transform is supported in  $[-X, X]$  takes exponential time in  $X$ , with the main contribution coming from hyperbolic terms. Because of the uncertainty principle, the factors mentioned above translate into an exponential running time in  $N$  (typically on the order of  $e^{cN^{1/3}}$ ). Moreover, the size of  $X$  required is very sensitive to the distribution of low eigenvalues; that is, an eigenvalue substantially smaller than average would greatly increase the required running time. This is likely the reason that we were unable to show Selberg's conjecture for conductor 857, for instance.

The bulk of this paper goes into addressing the first factor; separating forms by character and parity, sieving for newforms, and removing the contribution from special forms are all methods of thinning out the spectrum, so as to decrease the eigenvalue density. These techniques allowed us to broaden significantly the range of conductors that we could reach. A natural extension would be to use Hecke operators, which can further refine the spectrum at both small and large eigenvalues; we will pursue this further in [10].

Here we describe a more direct approach for dealing with the low eigenvalues. A prototype was given in the proof of Proposition 4.4; since the first ten eigenvalues for  $\Gamma(1)$  are known very precisely from [11], we could remove their contribution from the trace formula. A similar technique would work for testing Selberg's conjecture: For any fixed test function, the main contribution to the spectral side of the trace formula comes from roughly the first  $N$  eigenvalues. Thus, given a list of those eigenvalues, one could quickly verify the conjecture after removing their contribution. Moreover, as remarked in [11], the approach taken there for certifying eigenvalues can be generalized to  $\Gamma(N)$ , with polynomial complexity in  $N$  and the eigenvalue. Therefore, provided we have a fast method of locating eigenvalues, we get a polynomial-time algorithm for verifying Selberg's conjecture. Such a method is given heuristically by Hejhal's algorithms; see [36, 35] and the discussion of complexity therein. There is one subtle point to note: This approach still depends on finding Galois representations when they exist, since the method of [11] cannot distinguish eigenvalue  $\frac{1}{4}$  from  $\frac{1}{4} - \varepsilon$ . It's not hard to see that this part can be done in polynomial time as well, e.g. by a Hunter search. (Moreover, a targeted Hunter search used in conjunction with the trace formula is quite practical, as demonstrated in Section 5.) Thus, combining all of these

ingredients and *assuming Artin's conjecture*, there is an algorithm for verifying Selberg's conjecture that runs heuristically in polynomial time.

Finally, we mention a related question. In the proof of Proposition 4.4, we saw that the techniques of [11] may be combined with the trace formula to certify the first few eigenvalues on  $\Gamma(1)$ . By the analysis above, the same procedure should locate all eigenvalues up to a given  $R$ , heuristically in polynomial time in  $R$ . An appealing possibility is to remove the trace formula from this procedure, and certify the eigenvalues instead by analyzing the oscillations in the remainder term in Weyl's law, much like Turing's method for verifying the Riemann hypothesis. This approach will be taken up in a future publication.

## REFERENCES

- [1] A. O. L. Atkin and J. Lehner, Hecke operators on  $\Gamma_0(m)$ , *Math. Ann.* **185** (1970), 134–160.
- [2] A. O. L. Atkin and W. C. W. Li, Twists of newforms and pseudo-eigenvalues of  $W$ -operators, *Invent. Math.* **48** (1978), 221–243.
- [3] R. Aurich and F. Steiner, Staircase functions, spectral rigidity, and a rule for quantizing chaos, *Phys. Rev. A* (3) **45** (1992), 583–592.
- [4] E. B. Bogomolny, B. Georgeot, M.-J. Giannoni, C. Schmit, *Arithmetical Chaos*, *Physics Reports* **291** (1997), 219–324.
- [5] J. Bolte, C. Grosche, Selberg Trace Formula for Bordered Riemann Surfaces: Hyperbolic, Elliptic and Parabolic Conjugacy Classes, and Determinants of Maass-Laplacians, *Commun. Math. Phys.* **163** (1994), 217–244.
- [6] A. R. Booker, Artin's conjecture, Turing's method and the Riemann hypothesis, preprint, 2005.
- [7] A. R. Booker, Quadratic class numbers and character sums, to appear in *Math. Comp.*
- [8] A. R. Booker, A. Strömbergsson, web page containing data files and programs, <http://www.math.uu.se/~astrombe/stfz/>.
- [9] A. R. Booker, A. Strömbergsson, Galois representations of small conductor, in preparation.
- [10] A. R. Booker, F. Strömberg, A. Strömbergsson, Low eigenvalue statistics for congruence subgroups, in preparation.
- [11] A. R. Booker, A. Strömbergsson, A. Venkatesh, Effective computation of Maass cusp forms, preprint, 2005.
- [12] J.-M. Deshouillers and H. Iwaniec, Kloosterman sums and Fourier coefficients of cusp forms, *Inventiones Mathematicae*, **70** (1982), 219–288.
- [13] D. Doud and M. W. Moore, Even Icosahedral Galois Representations of Prime Conductor, to appear in *J. Number Theory*.
- [14] M. Eichler, *Lectures on Modular Correspondences*, Tata Institute, 1956.
- [15] S. Gelbart and H. Jacquet, A relation between automorphic representations of  $GL(2)$  and  $GL(3)$ , *Ann. scient. Éc. Norm. Sup.*, 4<sup>e</sup> série, **11** (1978), 471–542.
- [16] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, 1980.
- [17] D. A. Hejhal, *The Selberg Trace Formula for  $PSL(2, \mathbf{R})$* , Vol.1, *Lecture Notes in Math.* 548, Springer-Verlag, Berlin, 1976.
- [18] D. A. Hejhal, *The Selberg Trace Formula for  $PSL(2, \mathbf{R})$* , Vol.2 *Lecture Notes in Math.* 1001, Springer-Verlag, Berlin, 1983.
- [19] M. N. Huxley, Scattering Matrices for Congruence Subgroups, in *Modular Forms*, Ellis Horwood Series of Halsted Press, New York (1984), 141–156.
- [20] M. N. Huxley, Introduction to Kloostermania, pp. 217–306 in *Elementary and analytic theory of numbers*, Banach Center Publ. **17**, Warsaw, 1985.
- [21] M. N. Huxley, Exceptional eigenvalues and congruence subgroups, *Cont. Math.* **53** (1986), 341–349.
- [22] H. H. Kim and P. Sarnak, Refined estimates towards the Ramanujan and Selberg conjectures, appendix to H. H. Kim, Functoriality for the exterior square of  $GL_4$  and symmetric fourth of  $GL_2$ , *J. Amer. Math. Soc.*, **16** (2003), 139–183.



- [23] I. Kiming, Liftings of 2-dimensional projective Galois representations over  $\mathbb{Q}$ , *On Artin's conjecture for odd 2-dimensional representations*, Lecture Notes in Math., **1585**, Springer-Verlag, Berlin, 1994, pp. 8-36.
- [24] R. P. Langlands, Beyond endoscopy. In *Contributions to automorphic forms, geometry, and number theory*, pages 611–697. Johns Hopkins Univ. Press, Baltimore, MD, 2004.
- [25] R. P. Langlands, *Base change for  $GL(2)$* . Annals of Mathematics Studies **96**, Princeton University Press, 1980.
- [26] H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, *Math. Ann.* **121** (1949), 141–183.
- [27] C. Matthies and F. Steiner, Selberg's  $\zeta$  function and the quantization of chaos, *Phys. Rev. A* (3) **44** (1991), R7877–R7880.
- [28] T. Miyake, On automorphic forms on  $GL_2$  and Hecke operators, *Ann. Math.* **94** (1971), 174–189.
- [29] T. Miyake, *Modular Forms*, Springer-Verlag, New York, 1989.
- [30] J. Neukirch, *Algebraic Number Theory*, Springer-Verlag, 1999.
- [31] Number field tables available by ftp at [megrez.math.u-bordeaux.fr/pub/numberfields/](http://megrez.math.u-bordeaux.fr/pub/numberfields/)
- [32] A. Selberg, On the estimation of Fourier coefficients of modular forms, *Proc. Sypos. Pure Math.*, Amer. Math. Soc., vol. VIII, 1965, 1–15.
- [33] A. Selberg, Harmonic analysis, in *Collected Papers*, vol. 1, Springer-Verlag 1989.
- [34] J-P. Serre, Modular forms of weight one and Galois representations, in *Algebraic Number Fields*, A. Fröhlich (Academic Press 1977).
- [35] F. Strömberg, *Computational Aspects of Maass Waveforms*, Phd Thesis, Uppsala University, 2005.
- [36] F. Strömberg, Maass waveforms on  $(\Gamma_0(N), \chi)$  (computational aspects); to appear in Proceedings of the International School on Mathematical Aspects of Quantum Chaos II (Reisensburg, 2003).
- [37] A. Strömbergsson, The Selberg Trace Formula for Modular Correspondences (Licentiat thesis) U:U:D:M Report 1998:32. [<http://www.math.uu.se/~astrombe/papers.html>].
- [38] A. Strömbergsson, Some remarks on a spectral correspondence for Maass waveforms, *Internat. Math. Res. Notices*, 2001, No. 10, 505–517.
- [39] A. Strömbergsson, An application of an explicit trace formula to a well-known spectral correspondence on quaternion groups, 2000. [<http://www.math.uu.se/~astrombe/papers.html>].
- [40] A. Strömbergsson, The Selberg eigenvalue conjecture for squarefull conductor  $N$ , in preparation.
- [41] J. Tunnell, Artin's conjecture for representations of octahedral type. *Bull. Amer. Math. Soc.*, Vol. **5**, No. 2 (1981), 173–175.
- [42] A. B. Venkov, *Spectral Theory of Automorphic Functions*, *Proc. Math. Inst. Steklov* **153** (1982), ix+163 pp. (translation of *Trudy Mat. Inst. Steklov* **153** (1981)).
- [43] M.-F. Vignéras, *Arithmétique des Algèbres de Quaternions*, Lecture Notes in Math. 800, Springer-Verlag, Berlin, 1980.
- [44] M.-F. Vignéras, Représentations Galoisienne paires, *Glasgow Math. J.* **27** (1985), 223–237.